From the Institute of Microbiology and Epizootics
of the Department of Veterinary Medicine
at the Freie Universität Berlin
and
Helmholtz Institute for One Health (HIOH) Greifswald
and Robert Koch-Institute (RKI) Berlin

The Use of Fly iDNA to Monitor Wildlife and Study Sylvatic Anthrax Ecology at Human-Wildlife Interfaces

Inaugural-Dissertation
for the degree of
a Doctor of Philosophy (PhD) in Biomedical Sciences
at the
Freie Universität Berlin

submitted by
Mueena Jahan
Veterinarian from Dhaka, Bangladesh

Berlin 2023
Journal-No.: 4439

From the Institute of Microbiology and Epizootics of the Department
of Veterinary Medicine at the Freie Universität Berlin
and
Helmholtz Institute for One Health (HIOH) Greifswald
and Robert Koch-Institute (RKI) Berlin

The Use of Fly iDNA to Monitor Wildlife and Study Sylvatic Anthrax Ecology at Human-Wildlife Interfaces

---

Inaugural-Dissertation
for the degree of
a Doctor of Philosophy (PhD) in Biomedical Sciences
at the
Freie Universität Berlin

submitted by
Mueena Jahan
Veterinarian from Dhaka, Bangladesh

Berlin 2023
Journal-No.: 4439

Printed with permission
of the Department of Veterinary Medicine
at Freie Universität Berlin

Dean:	Univ.- Prof. Dr. Uwe Rösler
First reviewer:	Prof. Dr. Fabian Leendertz
Second reviewer:	Univ.-Prof. Dr. Marcus Fulde
Third reviewer:	Prof. Dr. Sébastien Calvignac-Spencer

Descriptors (according to CAB-Thesaurus):

Anthrax, *Bacillus cereus* biovar *anthracis*, Emerging infectious diseases, Mammals, Wildlife, Vectors, Monitoring ecosystems, Rain forests biodiversity

Promotion day: 01.12.2023

Table of Contents

List of abbreviations

AI	Artificial intelligence
ARUs	Autonomous recording units
Bcbva	*Bacillus cereus* biovar *anthracis*
BSL3	Biosafety level
CDV	Canine distemper virus
e.g.	Exempli gratia (for example)
eDNA	Environmental DNA
EIDs	Emerging infectious diseases
et al.	Et alia (and others)
HiWi	Hilfswissenschaftler (Student researcher)
HMPV	Human metapneumovirus
HRSV	Human respiratory syncytial virus
i.e.	It est (that is)
iDNA	Invertebrate derived DNA
PAM	Passive acoustic monitoring
PanAf	Pan African Programme
PCR	Polymerase chain reaction
PPE	Personal protective equipment
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus-2
SFV	Simian foamy virus
SIV	Simian immunodeficiency virus
SNP	Single nucleotide polymorphisms
STLV-1	Simian T-lymphotropic virus type 1
TNP	Taï National Park
UV	Ultra violet
WNS	White nose syndrome

1 Introduction

Global biodiversity has been declining rapidly due to massive anthropogenic pressure. These losses in biodiversity can eventually bring about major changes to the ecosystems themselves and to the ecosystem services they provide. Indeed, these changes in anthropogenic disturbance and biodiversity distributions have been associated with the emergence and re-emergence of pathogens in both human and wildlife populations (Devaux et al. 2019; Woolhouse 2002; Levins et al. 1994). The COVID-19 pandemic represents an extreme instance of an emerging infectious disease that can threaten many aspects of human livelihood and wellbeing, reminding us to consider disease emergence in our planning and our relationship with biodiversity (Piret and Boivin 2021; Schmeller et al. 2020). Most emerging infectious diseases are transmitted from animals to humans (hereafter referred to as zoonotic diseases). Of these zoonotic diseases, 72% are thought to have originated from wildlife (Butchart et al. 2010; Jones et al. 2008). To understand and ultimately prevent emerging infectious diseases, obtaining high-resolution information on wildlife species distributions as well as the respective diseases with zoonotic potential that they harbour, is critical. Such data could then form a cornerstone of One Health based disease mitigation efforts (Stadtländer 2015; Leendertz et al. 2006)

Systematic biomonitoring of both, wildlife and their infectious diseases is also very important in terms of conservation efforts, as infectious diseases have been linked to many wildlife population declines (Butchart et al. 2010). However, wildlife monitoring is a challenging task, particularly in tropical forests where biodiversity is highest and animals are difficult to observe due to their fear of people, a result of human hunting. Moreover, invasive sampling for pathogen detection or collection of genetic samples of host animals is a resource and time intensive procedure, which also comes with unavoidable risks to both the animals and the humans doing the sampling. Therefore, the development of non-invasive biodiversity monitoring tools is a critical area of active research.

Countries in the global south, including many in tropical Africa, Latin America and Asia are predicted to be at a higher risk of zoonotic disease emergence. This is thought to be because of the high biodiversity these countries harbour, in some cases coupled with poorly funded health care systems, making these areas emerging infectious disease hotspots. Paradoxically, these regions of the world with highest disease emergence risk generally have lower access to resources and a research infrastructure that could support cost effective monitoring tools (Jones et al. 2008).

As a resource and cost effective non-invasive approach, environmental DNA (eDNA) based biomonitoring has proven to be an effective and scalable tool in the field of wildlife and disease

monitoring studies (Stat et al. 2017; Valentini et al. 2016; Bohmann et al. 2014). A complex mixture of DNA from an environmental sample such as water, soil or animal excreta can be explored with metabarcoding approaches that allow researchers to determine the species diversity and sometimes also the pathogens in different ecosystems (Bohmann et al. 2014; Taberlet et al. 2012a). This monitoring system has proven particularly successful in revealing the biodiversity in marine aquatic environments (Valentini et al. 2016).

More recently, invertebrate derived DNA (iDNA) has also shown its promise for biomonitoring of terrestrial and tropical biodiversity (Hoffmann et al. 2016; Calvignac-Spencer et al. 2013b; Schnell et al. 2012). Invertebrates that come into close association with the host animal or their by-products as a part of their life cycle can take up the genetic material of animals. While there are many different sources of invertebrate-derived DNA (iDNA), carrion fly derived DNA based biomonitoring has shown particular potential for studying mammal diversity in terrestrial ecosystems. This is because flies can be so easily collected, are present in most ecosystems, and appear to preserve DNA for sufficient periods of time to allow it to be detected in random screening efforts (Gogarten et al. 2020; Hoffmann et al. 2018; Rodgers et al. 2017; Calvignac-Spencer et al. 2013a). Moreover, fly iDNA has also demonstrated its utility as a tool for detecting pathogens; indeed iDNA has shown its ability to reveal information about the ecology of a pathogen, even where the invertebrate sampler is not a vector of the pathogen under scrutiny (Bitome-Essono et al. 2017; Hoffmann et al. 2017). This iDNA based wildlife monitoring method is comparatively less resource and time intensive than traditional methods like censuses or camera traps, and may lend itself to biodiversity censuses and studies of disease ecology in precisely those regions of the global south that are poorly researched.

1.1 Aim of the thesis

The aim of this thesis was to assess and improve the possibility to use fly iDNA to generate data relevant to disease ecology, particularly in understudied rainforest ecosystems that represent disease emergence hotspots. The first chapter represents an in-depth exploration of an understudied pathogen's ecology using fly iDNA, which showcases the strengths and limits of iDNA. In this chapter, I studied the pathogen, *Bacillus cereus* biovar *anthracis* (Bcbva), responsible for sylvatic anthrax, in the tropical rainforest of Côte d'Ivoire, Taï National Park (TNP; Chapter 1). In my second chapter, given the strong evidence that flies frequently have contact with pathogens that could potentially emerge from biodiversity hotspots, I sought to explore the potential of flies to serve as a vector, moving pathogens from wildlife to humans and their livestock. Here I used a mark recapture experiment at the human-wildlife interfaces around Kibale National Park, Uganda, and explored the diversity of eukaryotic parasites found

on these flies (Chapter 2). Lastly, in Chapter 3, I explored the potential to optimize the extraction of fly iDNA to help make this method more scalable. My goal was to help make fly-derived iDNA biomonitoring more accessible to scientists working in regions of the global south.

This thesis represents the culmination of a vast sampling effort and extensive lab work. For Chapter 1, around 2,350 flies were trapped on a gradient spanning from the pristine forest within TNP to the surrounding villages, and I selected 500 of these flies and extracted them in 100 fly pools. For the work in Chapter 2, a total of 8,365 flies were captured and marked along the edges of Kibale National Park and 28,615 flies were examined to look for potential recapture events. Here too, extensive lab work was performed on the flies: I extracted the DNA from 1150 flies in 12 pools and sequenced the 18S rRNA of eukaryotic parasites. Lastly, for Chapter 3, as part of the Pan African Programme (PanAf), tens of thousands of flies were collected across the range of chimpanzees; a total of 784 flies were extracted from 8 different sampling sites throughout sub-Saharan Africa and previously analyzed in efforts to validate fly-derived iDNA approaches. For my thesis, I re-extracted a further 840 flies from this collection using two unexplored extraction methods.

2 Background

In this section, I provide information and critical background for the three studies in this thesis. First, I present an overview of wildlife population declines in general, factors related to these population declines, and how infectious diseases are involved in these wildlife population declines (section 2.1). In section 2.2, I explore aspects of wildlife population and disease monitoring and explain various key concepts for wildlife monitoring. Section 2.3 explores the potential role of fly derived iDNA for monitoring mammal species diversity and highlights how fly iDNA based pathogen surveillance can provide insights into the study of a pathogen's ecology.

2.1 Wildlife population declines

Wildlife population declines refer to the significant and persistent reduction in the abundance and distribution of wild animal species over time (Dirzo et al. 2014). The expeditious decline of wild populations across various ecosystems has appeared as a critical concern for the global scientific community and conservationists. As anthropogenic activities intensify, ecosystems face unprecedented pressures, leading to widespread disruptions in the delicate balance of nature. The alarming trend of wild population declines, characterized by dwindling numbers and shrinking habitats of different wild animal species, poses severe threats to the preservation of biodiversity, ecological stability, and the critical ecosystem services provided by the ecosystems that these species inhabit (Ogutu et al. 2011).

The global decline of biodiversity is a growing concern across the world. According to the living planet report released in 2022, which combined diverse monitoring efforts targeting wild mammals, fish, reptiles, birds and amphibians, the relative abundance of wildlife populations has plunged by an average of 69% since 1970, whereas the human population continues its booming upsurge, reaching nearly 8 billion. These changes appear to have the potential to further jeopardize the existence of many animals, including humans. An analysis of the global geographic representation of species populations declines showed that Latin America and the Caribbean have the fastest rates of decline, with around 94% wild vertebrate population abundance declines observed in the last nearly six decades, followed by Africa (66%), Asia and the Pacific (55%), North America (20%) and Europe and Central Asia (18%) (Living Planet Report, 2022). The situation is dire and highlights that there are rapidly changing distributions of animals in these ecosystems across the globe.

2.1.1 Causes of wildlife population decline

Anthropogenic disturbance plays a key role in local and global biodiversity losses. Anthropogenic disturbance refers to changes or disruptions to the natural environment caused by human activities. These disturbances include massive deforestation through exploiting natural habitats and reducing the availability of food and shelter for many species (Laurance et al. 2014; Wilcove et al. 2013, Hoekstra et al. 2005); poaching, illegal trading and overexploitation of wild animals, and climate change. The introduction of invasive species who can outcompete native species (Doherty et al. 2016) and environment pollution (Grier 1982) are also the consequences of anthropogenic disturbances that can negatively impact wildlife populations.

Collectively, these changes can have significant impacts on ecosystems and their inhabitants, leading to losses of biodiversity. Changes in biodiversity and the interactions of humans with wildlife have also been linked to the emergence and reemergence of infectious diseases. Emerging infectious diseases (EIDs) can newly emerge or may have already existed in a population before, but are extensively increasing in occurrences irrespective of geographical ranges (Daszak et al. 2000; Morse 1995). Disease patterns and dynamics have been constantly changing due to increased anthropogenic pressures and environmental changes, which facilitates more human-wildlife contact and the interconnected world then allows for the rapid spread of diseases at a global level (Schmeller et al. 2020). The rate of disease emergence appears to have been skyrocketing over the last few decades and represents a constant threat to global health and the health of economies (Schmeller et al. 2020; Brownlie et al. 2006). Emerging infectious diseases are mainly caused by zoonoses (60.3% of EIDs), i.e., by pathogens with an animal origin. Around 71.8% of the zoonotic events are caused by wildlife pathogens (Jones et al. 2008).

The risk of disease emergence from wildlife varies throughout the world. An oft-cited analysis revealed a substantial risk of wildlife zoonotic and vector borne EIDs in lower latitude countries like tropical Africa, Latin America and Asia. These are referred to as merging disease ‘”hotspots” and are typically found in areas where anthropogenic pressure is significant (Jones et al. 2008). There are certain human induced risk factors like human encroachment, deforestation, and habitat degradation through land expansion that are forcing wild animals to come closer into contact with humans and hence facilitate the likelihood of successful disease transmission events (Jones et al. 2008; Daszak et al. 2001). For instance, seemingly due to agricultural land expansion, a zoonotic pathogen, the nipah virus circulating in fruit bats in Malaysia, was able to enter into a commercial pig farm in the late 90s. From its natural host, the fruit bats, the virus spread to the pig farm and subsequently caused outbreaks in human

populations (Chua 2003). Human activities such as hunting, poaching, and illegal wildlife trade can also contribute to the spread of diseases by disrupting natural habitats and introducing pathogens to new areas. For example, several Ebola virus outbreaks have been linked to human hunting or scavenging, highlighting the potential for human activities and encroachment into remaining forests to contribute to the emergence of zoonotic diseases (Leroy et al. 2004).

Therefore, it is critical to understand the emergence of wildlife infectious diseases. At the same time, as the next section will highlight, changing disease dynamics and contact between humans and wildlife can be causes of wildlife population declines themselves, sometimes exacerbating already challenging situations for wildlife struggling in the face of anthropogenic disturbance.

2.1.2 Implications of diseases on wildlife populations

Emerging wildlife infectious diseases include a larger cohort of pathogens that affect terrestrial, freshwater and marine habitats (Daszak et al. 2001). Historically, wildlife diseases were only studied and considered as important when human and agriculture were threatened. But as the majority of emerging infectious diseases originate in wildlife, and wildlife disease outbreaks are also threatening the existence of endangered wild animals (Wright 2011), researchers have begun to study wildlife pathogens, which has revealed the impact of these pathogens on wildlife as well. Indeed, diseases are now thought to play a key role in many wildlife population declines and thus represent a major concern for the conservation of biodiversity as well.

Infectious diseases can spread rapidly among wildlife populations, leading to high mortality rates and population declines. This is the case, for example, with amphibian chytridiomycosis, a fungal disease that has been responsible for huge mortality and the decline of many amphibian species, including some that are critically endangered (Scheele et al. 2019; Daszak et al. 2001). Another example is white-nose syndrome (WNS), a fungal disease that has caused a dramatic decline in hibernating bat populations in North America in 2006. The disease affects bats while they are hibernating, causing them to wake up prematurely, which leads to depletion of their energy reserves, and ultimately death. Since its emergence, this disease has had a detrimental effect on the bat population and has been linked to up to 90% population declines for some bat species. This outbreak is considered one of the most severe wildlife diseases in North America (Cryan et al. 2010).

Another prominent example of a pathogen causing a major wildlife decline is sylvatic anthrax, caused by *Bacillus cereus* biovar *anthracis* (Bcbva). This pathogen chromosomally belongs to

the *Bacillus cereus* clade. However, its two virulence plasmids, PX01 and PX02, which are also found in *Bacillus anthracis*, implicated a pathogenic potential of Bcbva. Previous studies revealed the presence of Bcbva in animal carcasses, bones and carrion flies that were sampled extensively for more than three decades in TNP in Côte d'Ivoire (Hoffmann et al. 2017). These data suggested that nearly 40% of wildlife mortality in this rainforest ecosystem during the study period were due to this pathogen (Hoffmann et al. 2017). Since its first detection, this forest anthrax or sylvatic anthrax, has been found in a broad range of mammal hosts with a geographical distribution throughout West and Central Africa (Romero-Alvarez et al. 2020; Antonation et al. 2016). Unlike classical anthrax, which mainly affects ungulates, Bcbva has been found to cause mortality in a variety of wild mammal species, including non-human primates, ungulates, carnivores, rodents and bats in rainforest habitat (Hoffmann et al. 2017). Efforts to model the severe impact of forest anthrax in a biodiversity hotspot suggested that this pathogen will lead to the extirpation of local chimpanzee in TNP in the next 40 years, which raises significant concerns in terms of wildlife conservation (Hoffmann et al. 2017).

Wildlife infectious diseases can also affect top predators, which can in turn have cascading effects throughout the food web, altering the abundance and distribution of other species. For example, in the Serengeti ecosystem in Tanzania, African lions were infected by the Canine distemper virus (CDV), which led to the decline of the lion population by up to 30-40%. Due to the declining lion population, the number of other predator species, such as hyenas and leopards, appears to have increased in abundance and eventually altered the relative interactions within the predator-prey network (Nikolin et al. 2017). These examples demonstrate that studying wildlife infectious diseases can provide valuable insights into animal conservation as well.

2.2 Advances in wildlife monitoring and pathogen detection

In this section, I discuss the concept of wildlife monitoring in different ecosystems and how diverse monitoring tools can be used to detect the negative impact of anthropogenic pressure on existing wildlife in different terrestrial ecosystems. To this end, I highlight different traditional and newly emerged monitoring techniques and discuss limitations of these techniques. I argue that developments in eDNA monitoring, and iDNA specifically, could mitigate current gaps in wildlife monitoring studies, as well as studies of wildlife pathogens.

Wildlife monitoring can be broadly defined as the systematic observation and data collection of natural environments or any of their components for the analysis of wildlife species,

populations, and communities to gain insights into their abundance, distribution, behaviour, and ecological health (Zwerts et al. 2021). Systematic monitoring of both wildlife populations and their infectious diseases is key to influence conservation efforts and helps drawing an outline for the mitigation of disease emergence (Butchart et al. 2010). Large scale and long-term wildlife monitoring can play a vital role in combatting global biodiversity loss and contributing to data driven conservation efforts (Lindenmayer et al. 2022); (Calvignac-Spencer et al. 2013b). Monitoring helps not only assessing the effectiveness of management strategies of a particular ecosystem or its existing species communities and accelerate decision-making processes for the vulnerable and endangered species, but also tells us about the diseases related to particular communities (Ryser-Degiorgis 2013; Boadella et al. 2011)

Sampling of pathogens is another essential element for successful wildlife disease monitoring. Wild animals are typically sampled by a combination of invasive and non-invasive approaches to detect wild pathogens. Although sample quality in invasive sampling is generally higher, the invasive sampling of wild animals, especially endangered animals like wild great apes is a matter of ethical concern as it requires human-animal interaction and carries with it some risk to both humans and animals. Therefore, non-invasive techniques are often used as a sustainable alternative that allows the diagnosis of infections without causing additional stress to the wild animals (Köndgen et al. 2010). They typically target animal excreta (i.e., feces, urine, and saliva from food residues) that can be opportunistically collected. Non-invasive sampling has been widely used to detect gastrointestinal pathogens like adenoviruses, but even pathogens without a gastric tropism, such as simian immunodeficiency virus (SIV), simian foamy virus (SFV), and hepatitis B virus, have been detected in feces (Liu et al. 2008; Keele et al. 2009; Makuwa et al. 2005; Santiago et al. 2002). Feces samples can also be used to detect malaria (De Nys et al. 2013) and respiratory pathogens like human metapneumovirus (HMPV) and human respiratory syncytial virus (HRSV) from wild chimpanzees (Köndgen et al. 2010). Even antibodies can be detected non-invasively, as demonstrated for simian T-lymphotropic virus type 1 (STLV-1) antibodies in urine from wild chimpanzees in Taï National Park, Cote d'Ivoire (Leendertz et al. 2004) and Ebola virus antibodies from fecal samples of wild great apes in Gabon (Mombo et al. 2020). In another study, *Staphylococcus aureus* colonization in wild great apes and lemurs was identified using fecal and saliva samples (Schaumburg et al. 2013). Still, despite their promise, non-invasive approaches often require fresh samples collected from habituated primates, or expert trackers who can follow non-habituated animals at a distance and identify the species that gave rise to the sample.

Therefore, the selection of biodiversity monitoring tools is critical can be very challenging, as both, their efficiency and limitations need to be accounted for. Zwerts et al. (2021) outlined four criteria to consider when seeking to maximize the efficiency of wildlife monitoring efforts. The study suggest that wildlife surveillance should aim for 1) a specific targeted species coverage

(e.g., individual species or species community), 2) feasible population metrics (species density, species distribution or occupancy or encounter rate), 3) appropriate monitoring tools for species identification and 4) required resources and manpower for data collection and data analyses (Zwerts et al. 2021). Efforts to increase the efficiency of available monitoring tools and adapt new tools for a better understanding of animal ecology aim to reduce the required expertise and costs associated with their implementation.

2.2.1 Monitoring tools for wildlife surveillance

Different forms of wildlife monitoring tools can be applied based on different objectives. Direct observation of animals to estimate species abundance and host diversity is frequently used. There are different monitoring schemes for the direct monitoring of wild animals, such as line transects, capture-mark-recapture experiments or trapping of wild animals (Snaddon et al. 2013; Ahumada et al. 2011; Plumptre 2000). In areas where animals are hunted, such direct observation methods are challenging to implement as animals may reduce their noise and calling behaviour, as well as generally be extremely wary of people. In addition, these methods are often both, time consuming and require the expertise to identify animals or their signs or calls (Campbell et al. 2011). There are, of course, many benefits to such direct observation-based approaches; for example, the presence of research groups in an area helps to discourage poaching and has been shown to provide a protective effect for a protected area (Campbell et al. 2011). Such direct observation can also allow for the detection of disease outbreaks, for example when animals are visually observed with lesions or other signs of illness, as well the for the detection of carcasses (Kuisma et al. 2019; Hoffmann et al. 2017). Follow up investigations to collect invasive or non-invasive samples are frequently needed to confirm the observation-based diagnoses.

Considering the high manpower and expertise needed for implementation of such direct monitoring methods, these approaches have been increasingly augmented with remote sensing based monitoring, with tools such as camera or audio traps. Indeed, biodiversity assessments using camera trapping methods have been broadly used for terrestrial mammal communities (Gilbert et al. 2021; Wearn and Glover-kapfer 2019), vertebrates more broadly (Rovero et al. 2010; O'brien and Kinnaird 2008), reptiles (Richardson et al. 2018), and even arthropods using specialized cameras (Collett and Fishern 2017). Camera traps can also provide data on an animal species' activity, as well as on human behaviour in protected areas, such as poaching (Ramirez et al. 2021; Gaynor et al. 2018; Caravaggi et al. 2017).

Similar to direct observation-based surveys, camera trapping can also detect visible disease symptoms. For instance, leprosy in Cantanhez National Park, Guinea-Bissau, was detected

using a camera trap image (Hockings et al. 2021), as were cases of sarcoptic mange in wild boar in Switzerland (Ryser-Degiorgis 2013). However, the species detected and the encounter rate through camera trapping is highly influenced by the location where the device is installed. A major constraint to large scale camera trapping efforts has been the manual processing of large dataset of images, which is highly time intensive and requires expert validation. These limitations are gradually being overcome by the application of artificial intelligence (AI) algorithms which can annotate images automatically, as well as by the use of citizen science initiatives such as the platform Chimp and See (https://www.zooniverse.org/projects/sassydumbledore/chimp-and-see). However, such approaches are often not feasible for small scale camera trap studies, require high levels of logistic effort, are biased towards the detection of certain species, and their wide scale application is still limited (Duggan et al. 2021; Tabak et al. 2019; Norouzzadeh et al. 2018).

Audio trapping, also known as passive acoustic monitoring (PAM), has become a popular method for monitoring terrestrial biodiversity. This approach involves the use of autonomous recording units (ARUs) to capture wildlife sounds and vocalizations. These recordings are then analyzed to detect the presence of various species and assess their activity levels (Darras et al. 2019; Mielke and Zuberbühler 2013; Soltis et al. 2005). ARUs are effective in detecting a wide range of species, both large and small, based on their distinctive vocalizations or sounds (Wrege et al. 2017). This method is especially useful for capturing species that are difficult to detect through other means, such as insects (Ganchev and Potamitis 2007) and amphibians (Troudet et al. 2017; Aide et al. 2017).Bats are the most commonly monitored group using PAM, followed by birds and anuran amphibians (Sugai et al. 2019; Russo and Voigt 2016; Brauer et al. 2016; Brandes 2008). Both camera trapping and passive acoustic monitoring require resources in terms of costs and data processing. Camera trapping is often considered more expensive in terms of logistics, while setting up ARUs typically requires less training and is seen as a more cost-effective option (Zwerts et al. 2021).

DNA based capture-recapture is another approach that has been extensively used for estimating population abundance (Morin et al. 2016; Lampa et al. 2015; Taberlet et al. 1999) The DNA-based capture-recapture method is a combined approach of traditional capture-recapture studies (Amstrup et al. 2005) and genetic analysis aiming to provide information about the density of the species population, specifically the species that are difficult to capture or identify individually (Lukacs and Burnham 2005; Palsbøll et al. 1997). In this method, mainly non-invasive samples like hairs, feces and feathers are collected, DNA is extracted from the samples, and genetic tags or markers like microsatellites or single nucleotide polymorphisms (SNP) are selected and amplified using a polymerase chain reaction (PCR). The genotypes obtained from the genetic analysis are used to identify unique genetic profiles corresponding to individuals. This allows detections of the same genotypes to be considered recaptures,

indicating that they have been previously captured and identified (Palsbøll et al. 1997). This genetic tagging based monitoring technique has some limitations in terms of sampling and methodology. DNA-based capture-recapture methods depend on the availability of genetic markers specific to each species being monitored, that allow individuals to be distinguished. Developing species-specific markers can be time-consuming and the analyses are expensive. Besides that, there are often issues with amplification failures, allelic dropout and mutation or during genotyping or during amplification (Lukacs and Burnham 2005). As a result, while these methods are useful for certain projects, they have not become widely adopted for assessing the distribution and abundance of different species.

DNA based species identification has been taken a step further since the concept of DNA barcoding became included in the ecologist's toolbox (Taberlet et al. 2012b; Hebert et al. 2003). DNA barcodes are short, standardized DNA sequences that can be used as markers for the taxonomic identification of the organisms from which the DNA originated (Taberlet et al. 2018; Hebert et al. 2003). The standard barcoding approach is to identify species from more or less intact DNA isolated from an individual sample using PCR amplification, Sanger sequencing that amplicon, and comparing the obtained sequences with existing reference database of homologous sequences. More information to disentangle close relatives is encoded in variable sections of an amplified region, so regions are typically identified as suitable barcoding regions, when they contain sufficient variability to distinguish species at the level needed to address the biological question of interest. This means that rather than just focusing on the length of targeted DNA segments and the non-variability of the primer sites, during primer design, researchers typically look for variable regions flanked by conserved regions where a primer can bind (Taberlet et al. 2012b). A barcode thus typically consists of a highly variable short DNA region that is distinct for all species of the target taxonomic group, flanked by two conserved regions that act as primer binding sites during PCR reaction for unbiased amplification and preventing the amplification of this barcode for undesirable taxa (Taberlet et al. 2018). Traditional DNA barcoding is limited to working with relatively pure samples that contain only a single taxon that would be amplified with the selected primers. In addition, Sanger sequencing is quite cost ineffective, suggesting it might not be the most suitable approach for efficient high-throughput species identification in ecological studies (Taberlet et al. 2012b). Several factors must be carefully considered while selecting a barcoding region, such as that the targeted taxonomic group must be explicitly defined and the primers must preferentially target these organisms.

Barcoding techniques have been scaled up with the advent of next generation sequencing technologies. This allows for the simultaneous identification of many barcodes in a single mixed sample containing many species that are amplified with the specific particular primer set, an approach that has been termed as metabarcoding (Taberlet et al. 2012a). Like

barcoding, it involves the targeted PCR amplification and sequencing of specific DNA regions, but the diversity of amplicons is sequenced with next generation sequencing technologies, and each of these barcodes can then be assigned to a taxonomic rank. Because samples targeted with metabarcoding often contain a diversity of species, the ideal primer set will not amplify the DNA from non-target organisms. When this is not possible, amplification of non-target species can be further enhanced by including blocking oligonucleotides in so called blocking primers; these can be designed so that they specifically bind to the specific sequence of a non-target group and block or significantly reduce the amplification of the sequence from these taxa. In addition, because the DNA quality targeted with metabarcoding often varies, another important consideration is the size of the metabarcoding region targeted; it should correspond to the level of target DNA degradation. For instance, a shorter metabarcoding region (<100-150 bp) may work more efficiently for use with highly degraded DNA (i.e. eDNA), while relatively longer metabarcodes (<250-300bp) can be used for either higher quality DNA or for samples where there is a vast surplus of target species DNA (e.g., bacterial microbiome studies (Taberlet et al. 2018).

A key goal of DNA metabarcoding is extracting the taxonomic composition of environmental samples using sequence data. Taxonomic annotation of DNA sequences primarily means comparing an unidentified eDNA sequence with a high quality metabarcode reference database that contains taxonomic information (Taberlet et al. 2018). When comparing an unknown sequence to a reference database, researchers can determine the most likely taxonomic category or species to which it belongs. Reference databases thus facilitate the evaluation of biodiversity within environmental samples, and as these database are increasing in quality and coverage, old datasets can be reanalysed to reveal previously undetected or taxonomically unassignable biodiversity. Through the comparison of DNA sequences with the database, researchers can estimate the range of species existing in a specific sample. This holds significant importance for ecological investigations, conservation initiatives, and the comprehension of how ecosystem dynamics function (Taberlet et al. 2012b). Even when no taxonomic reference databases are available to assign a sequence to a specific taxonomic rank, researchers can use so called molecular operational taxonomic units to create assignments of sequences to groups. In addition, because sequence data is generated for each assignment, phylogenetic relationships can sometimes be quite accurately inferred from the barcoding sequences themselves.

2.2.2 Application of eDNA for wildlife surveillance and pathogen detection

Environmental DNA (eDNA) based biomonitoring is the analysis of genomic DNA from different organisms shed into the environment; for example, in soil, sediments, fresh water, or even in the air (Taberlet et al. 2012b). DNA metabarcoding now allows for the identification of multiple taxa from a single environmental sample at a time and has successfully been applied in the field of ecology for exploring community diversity and for detecting rare and elusive taxa, as well as for identifying pathogens in ecosystems (Bohmann et al. 2014). eDNA has been extensively applied as a biodiversity monitoring tool, with a focus on identification of both broad range organisms. Metabarcoding of eDNA has proven a cost and time effective solution to detect rare, endangered and cryptic species present at a lower abundance and difficult to identify with direct observation techniques (Stat et al. 2019; Piaggio et al. 2014; Jerde et al. 2011). At the same time, a combination of molecular approaches, such as microsatellites or hybridization capture to sequence genomic regions of interest, have allowed explorations of genetic diversity of specific animals from environmental samples (Danabalan et al. 2023; Adams et al. 2019).

eDNA based monitoring has already proven effective in aquatic ecosystems to survey mammal and invertebrate diversity, as large amount of vertebrate DNA can be dispersed through broader waterbodies (Valentini et al. 2016). eDNA based monitoring of freshwater primarily focuses on detecting individual species (Goldberg et al. 2011; Jerde et al. 2011). However, it also aims to identify diverse taxa using metabarcoding (Thomsen et al. 2012a). In addition, macrofauna and meiofauna from marine ecosystems were also explored through eDNA analysis (Thomsen al. 2012a). These tools have been integrated into management efforts, to monitor the spread of invasive species and changes in the distribution of key species in marine ecosystems (Thomsen et al. 2012b). Several studies also showed that environmental samples like soil (Andersen et al. 2012) or water from watering holes (Rodgers and Mock 2015) can contain amplifiable terrestrial vertebrate DNA. However, as yet, the utilization of eDNA in terrestrial biodiversity monitoring is not as significant as in marine ecology. This phenomenon has been attributed to several factors such as that large volume soil analyses for vertebrate diversity entail extensive costs and logistic difficulties, as eDNA tends to attach to clay particles or organic matter (Taberlet et al. 2018). Besides that, research has demonstrated that plant DNA endures in soils for several centuries (Yoccoz et al. 2012), whereas, under favourable conditions, vertebrate DNA has been observed to remain detectable in soil for multiple years. While this enables the reconstruction of past biodiversity and provides researchers with a way to bypass temporal or seasonal fluctuations that impact other survey techniques, it may hinder

the differentiation between contemporary and historical signals (Taberlet et al. 2018; Andersen et al. 2012). In addition, it has been suggested that the effectiveness of eDNA might be diminished in tropical zones due to the high biodiversity, which could hinder the application of universal primers capable of providing satisfactory taxonomic precision for the comprehensive identification of all biodiversity and high temperatures and humidity could accelerate the degradation of DNA, thereby introducing a factor that might contribute to high occurrences of false negatives and hampering taxonomic identification (Carvalho et al. 2022).

Most eDNA studies have focused on individual species detection (Ficetola et al. 2008) or assessment of species diversity on a large scale (Bohmann et al. 2014; Rees et al. 2014). Indeed, applying eDNA-based monitoring to study host abundance and host occupancy has faced challenges. Several factors may be involved in these challenges; one stems from the fact that eDNA does not detect genetic variance within the species, hence cannot accurately determine species abundance of hosts. Moreover, host occupancy studies using eDNA might be impeded by the potential persistence of DNA in the environment (Andres et al. 2023). DNA can persist in the surroundings for varying durations, leading to false indications of host presence even when the hosts have moved away from the monitored site (Farrell et al. 2021). In addition, the effectiveness of eDNA based techniques could be hindered by limited taxonomic resolution. Universal primers used for amplification may not provide sufficient specificity to accurately identify all host species or strains present, particularly in areas with high biodiversity (Taberlet et al. 2018). On the other hand, genomic invertebrate DNA has proven a cost effective tool for estimating the distribution of different vertebrate species throughout entire protected areas, while covering a wide range of taxonomic groups (Ji et al. 2022) and thus facilitating direct measurement of biodiversity conservation outcomes.

Beyond assessments of animal biodiversity, eDNA has also been increasingly employed for the identification, investigation and surveillance of pathogens (Farrell et al. 2021; Bass et al. 2015). Non-invasive genomic DNA from environmental samples can be used to identify known pathogens in an area, but can also reveal information about other uncharacterized members of pathogenic lineages, thus making eDNA a useful tool for exploring a pathogen's distribution and genetic diversity (Alfano et al. 2021). Metabarcoding of eDNA samples has also been proven useful for detecting pathogens and parasites, mostly from aquatic ecosystems (Stat et al. 2017; Bass et al. 2015). While such eDNA biomonitoring has identified pathogens (Huggins et al. 2018) and eukaryotic parasites (Seeber et al. 2017) from terrestrial ecosystems as well, its application for terrestrial pathogen monitoring and surveillance is still in its infancy.

Although eDNA based biodiversity monitoring offers numerous advantages, there are potential contamination risks associated with the process. Since samples are collected from the environment, the DNA concentration is often relatively low and can be mixed with other

contaminants that inhibit PCRs. eDNA may also be more susceptible to DNA degradation since it might have been exposed to environmental influences for extended period of time, which can also interfere with analyses and interpretation of results (Bohmann et al. 2013).This is particularly true in habitats like tropical rainforests, where samples quickly decompose and also non-invasive samples often contain poor quality nucleic acids which inhibit PCR and can interrupt the proper analysis (King et al. 2015). Including positive and negative controls during PCR amplification could be an effective solution to monitor both, contamination and reaction success. Moreover, maintenance of clean and sterile laboratory environments by wearing personal protective equipment (PPE) and applying physical barriers like clean work benches or laminar air flow cabinets are the prerequisite for reducing the risk of contaminating samples with extraneous DNA and to prevent airborne contaminations.

2.2.3 Application of invertebrate DNA for host and pathogen detection in terrestrial ecosystems

To better understand terrestrial host population diversity, researchers in the field of eDNA based biomonitoring have begun exploring the analysis of invertebrate genomic DNA as a cost effective alternative tool (Rodgers et al. 2017; Hoffmann et al. 2016; Calvignac-Spencer et al. 2013a; Calvignac-Spencer et al. 2013b). Indeed, invertebrate derived DNA (iDNA) has shown promise as a tool for terrestrial biodiversity monitoring. This field of research grew out of the field of human and livestock disease epidemiology, where genotyping of host DNA in hematophagous invertebrates was used to explore host-pathogen links and the role of vectors in parasite transmission (Kent 2009). Invertebrates that depend on either vertebrate hosts or their carcasses as a nutritional source or to complete their lifecycle are often considered ideal candidates for iDNA based monitoring. Particularly those invertebrates that can be easily captured in the field and that preserve their hosts DNA well are ideally suited for iDNA efforts (Calvignac-Spencer et al. 2013a).

In tropical and temperate ecosystems, both hematophagous and non-hematophagous insects like mosquitoes, leeches, sand flies, and carrion flies (flies that feed on animal carcasses) have been effectively used for mammal biodiversity assessments (Rodgers et al. 2017; Hoffmann et al. 2016; Grubaugh et al. 2015; Schubert et al. 2015; Calvignac-Spencer et al. 2013a; Schnell et al. 2012). Studies have demonstrated the potential of iDNA to monitor local mammal communities, as well as for the targeted detection of specific species of interest (Schubert et al. 2015; Calvignac-Spencer et al. 2013b). iDNA has also demonstrated its usefulness in

detecting diverse mammal species, including small, elusive animals, as well as those living at a low density (Bohmann et al. 2013). Furthermore, iDNA can also contribute information about a host species' abundance and occupancy on a large geographic scale. Ji et al. (2022) introduced leech iDNA metabarcoding as a promising tool for gauging spatially resolved vertebrate occupancies across entire protected areas and with a comprehensive representation across a wide range of taxonomic groups (Ji et al. 2022).

Beyond detecting vertebrates, iDNA has also shown potential to detect pathogens in both, wildlife and human communities, contributing information about the genetic diversity and distribution of known and hitherto uncharacterized wild pathogens (Hoffmann et al. 2017; Grubaugh et al. 2015). Time series have leveraged iDNA to explore pathogens' spatio-temporal dynamics as well (Hoffmann et al. 2017).

2.3 Fly iDNA based biomonitoring

In the rest of this thesis, I refer to invertebrates belonging to three non-blood sucking dipteran fly families, namely Calliphoridae (blow flies), Muscidae (house flies) and Sarcophagidae (flesh flies), all are commonly known as carrion flies, simply as flies. Carrion flies tend to colonize fresh carcass so they are also called "early colonizers" (Richards et al. 2013). These flies feed on carcasses as a part of their lifecycle and female flies and their larvae feed on fresh carcass as high nutritive source during their oviposition and larvaposition. Flies are extremely abundant in most terrestrial ecosystems and their close association with a diversity of mammal species predestines them for mammalian biodiversity assessments. Besides that, fly trapping methods and the baits used for fly trapping are significantly inexpensive and can be reused and flies can be preserved at room temperature on silica gel, require no electricity for cooling samples in the field . Moreover, the training of field assistants for fly trapping is a simple process and fly sampling effort can easily be increased by applying multiple traps in a single location therefore this approach can be potentially cost effective (Gogarten et al. 2020).

Many studies have explored fly-derived iDNA as a tool for the rapid assessment of mammal diversity in tropical and temperate ecosystems (Rodgers et al. 2017; Hoffmann et al. 2016; Schubert et al. 2015; Calvignac-Spencer et al. 2013b). At the same time, they have shown that fly DNA can be used to detect pathogens (Gogarten et al. 2019; Hoffmann et al. 2017). This body of literature suggests fly iDNA may now be poised to be able to make important contributions to understanding tropical disease ecology. I discuss and develop these ideas in the following sections.

2.3.1 Mammal diversity monitoring using fly derived DNA

Knowledge about animal distributions is crucial to infer alterations in wild populations and aid in conservation efforts. Flies are often closely associated with mammal or vertebrate hosts and take up genetic material either as a meal source or mechanically on their body parts when they come into direct contact with the mammal carcass, the animals themselves, or their by-products, thus suggesting that flies have the potential to represent useful sources of iDNA for exploring mammal biodiversity in tropical terrestrial ecosystems. Metabarcoding enables fly iDNA exploration in a rapid and cost effective manner to explore wild host species distributions.

Fly iDNA based biomonitoring has been successfully employed for assessing wild mammal diversity with a broad range of body sizes in tropical habitats of Côte d'Ivoire and Madagascar (Calvignac-Spencer et al. 2013b). Earlier research showed that targeted mammal species detection using fly derived iDNA could increase the species detection ability compared to other more broad taxonomic detection approaches (Schubert et al. 2015). Direct comparisons of fly iDNA with camera trapping demonstrated that even a small collection of flies had the ability to provide useful data on mammal species community composition in both, arid and rainforest ecosystems across sub-Saharan Africa (Gogarten et al. 2020; Calvignac-Spencer et al. 2013b). The work by Gogarten et al. (2020) demonstrated that flies are able to detect very small rodents and bats that can easily be missed by camera trapping efforts or direct observation. Such fly iDNA tools also showed potential for monitoring urban mammal species in a temperate environment (Hoffmann et al. 2018). Fly iDNA has even recently been effectively used to reconstruct mammal mitogenomes, opening the exciting possibility of using fly iDNA to study population level genetic diversity (Danabalan et al. 2023).

2.3.2 Detection of pathogens using fly derived DNA

The detection and monitoring of wildlife pathogens is challenging, particularly inside dense, biodiverse tropical forests, and if one wishes to sample without interrupting wildlife activities. Nevertheless, fly iDNA has been demonstrated as suitable to detect some wildlife pathogens (Hoffmann et al. 2017).The rationale behind using the fly as a potential candidate for pathogen detection is that flies can take up and transmit genetic material of host pathogens both, mechanically or via ingestion and faecal deposition while coming into contact with wild animals,

their excreta and carcasses (Calvignac-Spencer et al. 2013a; Förster et al. 2007). Flies might thus be hypothesized to be particularly adept at sampling pathogens that occur in high copy numbers in feces (e.g., gastrointestinal pathogens) or in carcasses (e.g., anthrax). Indeed, flies have been implicated in the transmission of different pathogens and parasites. For example, studies conducted in the 1930's under laboratory conditions implicated flies as a potential vector involved in the transmission of *Treponema pallidum* subsp. *pertenue*, the pathogen responsible for human yaws disease. These studies showed that flies which fed on the lesions of a person could then transmit the pathogen to rabbits. Based on these observations, Knauf et al. (2016) explored the potential for flies as a useful tool for pathogen detection in natural ecosystems. Indeed, the detection of *Treponema pallidum* subsp. p*ertenue* DNA in flies in an ecosystem where non-human primates were known to be infected demonstrated the power of flies for pathogen surveillance. In regions with high occurrence of *Treponema* infections, flies might also act as vectors involved in the transmission of yaws among humans or non-human primate populations, though further experiments are needed to confirm their role in transmission in natural settings.

Several studies have demonstrated that different types of pathogens can be detected within flies, suggesting that fly iDNA based surveillance might contribute to multiple dimensions of wild pathogen monitoring. For instance, fly iDNA has been used to detect genomic DNA of various mammal pathogens like, adenoviruses (Hoffmann et al. 2016); Bcbva (Hoffmann et al., 2017); and *Treponema pallidum pertenue* (Gogarten et al. 2019) in tropical rainforest ecosystem in western Africa where these pathogens were known to circulate. Fly iDNA also was used as unbiased sampler of emerging pathogen for which the host range and geographic distribution were initially poorly understood or changing rapidly. Hoffmann et al. (2017) used fly iDNA, along with other samples from carcasses and bones to provide a seminal study of the epidemiology of this novel pathogen. Hoffmann et al. (2017) were able to identify a broader range of mammal species in flies that carried this pathogen that inferred from carcass and bones which is a clear indication that fly iDNA based wildlife pathogen monitoring can contribute to understanding the ecology and epidemiology of a pathogen. Fly iDNA based pathogen surveillance can also help in understanding the distribution pattern of the pathogen at various spatial scales by detecting pathogen prevalence in both known and unknown sites of occurrence. In their study, Hoffmann et al. (2017) investigated the prevalence of Bcbva within and around the research area of TNP using fly iDNA analysis. Their research revealed the capability of fly iDNA to expand the known geographic distribution beyond TNP to Liberia in western Africa, where Bcbva had not been previously detected. At the same time, Gogarten et al. (2019) showed that flies seem to form an association with primate social groups which can last for several weeks; flies trapped in particular primate social groups revealed pathogens infecting those particular monkeys, with whole genome sequencing of Bcbva from a fly being

epidemiologically linked to the isolate that killed a monkey in the group a week prior. This study suggests fly iDNA can tell us about the pathogens circulating in a particular monkey group as well. Overall, fly iDNA shows potential to helping understanding a pathogen's ecology and geographic distribution.

Along with host diversity monitoring and pathogen detection ability, flies can also play a role in transmitting and disseminating some pathogens. Carrion fly possess several unique traits that explicitly indicates fly prospects to act as a vector for wildlife disease transmission. Flies have been found to take rest in the vicinity of the food source after feeding and they also regurgitate to facilitate digestion and defecate in surroundings which lead to disseminate pathogens (Braak and de Vos 1990). A study found that nonhuman primate social groups retain higher fly densities inside their groups and this was driven in part by flies actually maintaining an association with a particular social group (Gogarten et al. 2022; Gogarten et al. 2019). Flies remained with a social group of terrestrial sooty mangabeys (*Cercocebus atys*) for up to 13 days and over a large distance (up to 1.3 km) (Gogarten et al. 2022; Gogarten et al. 2019). Given that these high density of flies also carried viable Bcbva (Gogarten et al. 2019), and that larger social groups of primates harbor higher fly densities, this suggests flies might increase the disease risk for primates living in larger groups. Fly iDNA can also aid in indirect transmission of viable zoonotic pathogens like monkey pox virus in rainforest ecosystems (Patrono et al. 2020). Indeed, for humans, increasing fly densities have been shown to increase human disease risk (Förster et al. 2007; Banjo et al. 2005; Graczyk et al. 2001; Greenberg 1971).

2.3.3 Exploring further potential of flies for studying disease ecology and optimizing fly iDNA based biomonitoring

As wildlife diseases are mostly uncharacterized and pose a risk for both wildlife and human health, the regular detection and monitoring of wildlife pathogens and an understanding about the epidemiology and ecology of emerging pathogens is critical (Alfano et al. 2021). As I mentioned in the previous section (2.3.2), fly iDNA has already shown promise as a tool for studying different aspects of wildlife pathogen diversity. One of the outstanding examples of the usefulness of fly iDNA was the identification and monitoring of epidemiology and distribution pattern of a novel pathogen, Bcbva at a broad geographic scale in sub-Saharan Africa (Hoffmann et al. 2017), while also revealing a higher prevalence of Bcbva in the research activity zone in TNP. A recent study suggested that humans living around TNP were

serologically exposed to Bcbva (Dupke et al. 2020) and a critical question is whether Bcbva is restricted to the rainforest ecosystems and humans are exposed on entering the park, or whether the pathogen also circulates in the anthropogenically used areas. In chapter 1, I address this question by exploring the distribution of Bcbva across a gradient spanning the pristine forest within TNP to the surrounding villages with fly iDNA. I worked at a much smaller geographic scale than previous studies to explore the distribution of Bcbva and using the same iDNA, while also seeking to reconstruct fly and mammal communities with metabarcoding to investigate potential links with Bcbva detection and changes in these communities across the gradient. I also examined the power of iDNA to contribute to genomic surveillance efforts of Bcbva by generating the whole genome of detected isolates and nesting this diversity in known Bcbva diversity.

As flies may also pose a risk of transmitting wildlife pathogens, a critical question is whether flies have a stable and close association with wildlife and rainforest ecosystems, or whether they may move outside of these ecosystems into areas used by humans and potentially serve as a vector of wildlife pathogens into these areas. In chapter 2, I illustrate how I investigated fly movement of non-human primate-associated flies across landscapes surrounding Kibale National Park, Uganda, using a mark-recapture experiment and exploring the pathogens that these flies carry. I elaborate on the role of flies as a potential vector between wild nonhuman primates, livestock and humans at this human-wildlife interface using fly iDNA.

Finally, metabarcoding of fly iDNA has shown immense potential as a cost effective terrestrial wildlife monitoring tool. Traditional fly iDNA extraction methods are time and cost intensive, precluding their application in resource poor settings. Therefore, in chapter 3 I explore the optimization of fly extraction methods with the goal of making fly iDNA based biomonitoring easier, cheaper and faster to scale up. In this chapter, I introduce two new extractions methods for analysing fly iDNA collected from eight different sampling sites throughout sub-Saharan Africa, which had been analysed previously with an established fly extraction method. I pursue the question whether two short cuts, the mass pooling of only one fly leg from each of a larger number of flies, or the bulk extraction of flies with non-digestive buffer methods, could simplify the traditional fly extraction method with intensified outcome.

3 Publications:

3.1 Publication I

Jahan, M., Lagostina, L., Gräßle, T. *et al.*

Fly iDNA suggests strict reliance of the causative agent of sylvatic anthrax on rainforest ecosystems

Published in Special Issue, *Environmental DNA* (2023)

Received: 04.11.2022

Accepted: 30.01.2023

Published: 15.03.2023

https://doi.org/10.1002/edn3.401

Received: 4 November 2022 | Revised: 30 January 2023 | Accepted: 17 February 2023

DOI: 10.1002/edn3.401

Environmental DNA
Dedicated to the study and use of environmental DNA for basic and applied sciences
Open Access
WILEY

SPECIAL ISSUE ARTICLE

Fly iDNA suggests strict reliance of the causative agent of sylvatic anthrax on rainforest ecosystems

Mueena Jahan[1,2,3] | Lorenzo Lagostina[1,2] | Tobias Gräßle[1,2] | Emmanuel Couacy-Hymann[4] | Leonce Kouadio[4,5] | Valère K. Kouakou[4] | Hermann A. Krou[4] | Arsène M. Mossoun[4] | Livia V. Patrono[1,2,6] | Kamilla Pléh[1,2] | Jonas A. Steiner[1] | Noma Yves[7] | Fabian H. Leendertz[1,2] | Sébastien Calvignac-Spencer[2,6] | Jan F. Gogarten[1,2,6,8]

[1]Helmholtz Institute for One Health, Greifswald, Germany

[2]Epidemiology of Highly Pathogenic Organisms, Robert Koch Institute, Berlin, Germany

[3]Department of Microbiology and Public Health, Bangabandhu Sheikh Mujibur Rahman Agricultural University, Gazipur, Bangladesh

[4]Laboratory for Innovative Research for Emerging Diseases and Diagnostics, Centre National de Recherche Agronomique, NGO One Health for All, Abidjan, Côte d'Ivoire

[5]Université Peleforo Gon Coulibaly, Korhogo, Côte d'Ivoire

[6]Viral Evolution, Robert Koch Institute, Berlin, Germany

[7]Laboratoire d'écologie moléculaire et de génétique evolutive, Department of Genetics, University Felix Houphouët Boigny, Abidjan, Côte d'Ivoire

[8]Applied Zoology and Nature Conservation, University of Greifswald, Greifswald, Germany

Correspondence
Sébastien Calvignac-Spencer and Jan F. Gogarten, Epidemiology of Highly Pathogenic Organisms, Robert Koch Institute, Berlin, Germany and the Helmholtz Institute for One Health, Greifswald, Germany.
Email: calvignacs@gmail.com and jan.gogarten@gmail.com

Funding information
Deutsche Forschungsgemeinschaft; European Commission; Ministry of Science and Technology, Government of the People's Republic of Bangladesh

Abstract

Metabarcoding of invertebrate-derived DNA (iDNA) is increasingly used to describe vertebrate diversity in terrestrial ecosystems. Fly iDNA has also shown potential as a tool for detecting pathogens. Combining these approaches makes fly iDNA a promising tool for understanding the ecology and distribution of novel pathogens or emerging infectious diseases. Here, we use fly iDNA to explore the geographic distribution of *Bacillus cereus* biovar *anthracis* (Bcbva) along a gradient from the forest within Taï National Park, Côte d'Ivoire, out to surrounding villages. We tested fly pools ($N = 100$ pools of 5 flies) collected in the forest ($N = 25$ pools), along the forest edge ($N = 50$ pools), and near surrounding villages ($N = 25$ pools) for Bcbva. Using the same iDNA, we sought to reconstruct fly and mammal communities with metabarcoding, with the aim of investigating potential links with Bcbva detection. We detected Bcbva in 5/100 fly pools and positivity varied significantly across the habitat types (forest = 4/25, edge = 1/50, village = 0/25). It was possible to culture Bcbva from all positive fly pools, confirming their positivity, while sequencing of their whole genomes revealed a considerable portion of known genomic diversity for this pathogen. iDNA generated data about the mammal and fly communities in these habitats, revealing the highest

Sébastien Calvignac-Spencer and Jan F. Gogarten should be considered joint senior author.

mammal diversity in the forest and considerable changes in fly community composition along the gradient. Bcbva host range estimates from fly iDNA were largely identical to the results of long-term carcass monitoring efforts in the region. We show that fly iDNA can generate data on the geographic distribution and host range of a pathogen at kilometer scales, as well as reveal the pathogen's phylogenetic diversity. Our results highlight the power of fly iDNA for mammal biomonitoring and pathogen surveillance.

KEYWORDS
Anthropized environments, *Bacillus cereus* biovar *anthracis*, biodiversity, environmental DNA (eDNA), invertebrate-derived DNA (iDNA), mammals, sylvatic anthrax, wildlife infectious diseases

1 | INTRODUCTION

Anthropogenic disturbance is causing major declines in local and global biodiversity. These changes have been linked to the emergence of pathogens, not just in human populations, but among wildlife as well (Hale et al., 2022; Stegen et al., 2017). Systematic monitoring of both wildlife populations and their infectious diseases is needed to help inform conservation efforts and design disease emergence mitigation strategies (Butchart et al., 2010).

Long-term wildlife monitoring can be an effective tool for understanding mammal populations and their disease dynamics (Hoffmann et al., 2017; Kuisma et al., 2019). Unfortunately, long-term behavioral observations, that can lead to the recording of potential ill-health manifestations, and the detection and sampling of carcasses in the wild, are extremely resource intensive and, therefore, not particularly scalable. Camera trapping represents a means to detect a diversity of mammal species in an area, potentially even providing data on abundance (Gilbert et al., 2021), but can only detect extremely visible disease symptoms (Hockings et al., 2021) and cannot result in firm diagnosis, that is, the identification of the causative agent. Non-invasive sampling of mammal excreta (e.g., of feces, urine) has proven useful for detection of both mammals, the antibodies they produce when exposed to a pathogen, and the direct detection of pathogen genetic material (Mombo et al., 2020; Morin et al., 2016; Santiago et al., 2002), but fresh fecal sample collection can be challenging. This is particularly true in habitats like tropical rainforests, where feces quickly decompose and mammals that are not habituated to humans are typically hard to observe even briefly, let alone long enough to observe them defecating.

In contrast, trace amounts of animal DNA shed into the environment (environmental DNA, eDNA) can be detected with metabarcoding techniques and represents an effective and scalable tool for tracking the distribution of wildlife and at least some pathogens, particularly in aquatic ecosystems (Andruszkiewicz et al., 2017; Bohmann et al., 2014; Ficetola et al., 2008; Lynggaard et al., 2022; Stat et al., 2017). In terrestrial ecosystems, the animal DNA found in invertebrates that come into contact with animals or their by-products as part of their life cycle (e.g., leeches, sand flies, carrion flies) have shown their ability to rapidly assess mammal biodiversity in tropical and temperate ecosystems (Calvignac-Spencer et al., 2013; Gogarten et al., 2020; Hoffmann et al., 2018; Rodgers et al., 2017; Schnell et al., 2012). Crucially, invertebrate-derived DNA (iDNA) has also demonstrated its utility as a tool for detecting pathogens. Thus, iDNA can reveal information about the ecology of a pathogen, even where the invertebrate sampler is not a vector of the pathogen under scrutiny (Bitome-Essono et al., 2017; Hoffmann et al., 2017).

Such iDNA-based pathogen surveillance can be particularly useful for novel pathogens or emerging infectious diseases, for which the host range and geographic distribution are initially poorly understood or rapidly changing. For example, a new form of anthrax caused by *Bacillus cereus* biovar *anthracis* (Bcbva) was identified in dead wild chimpanzees in a tropical rainforest ecosystem in Taï National Park (TNP), Côte d'Ivoire, in 2001 (Leendertz et al., 2004). Genomic studies revealed that while chromosomally distinct from the causative agent of savannah anthrax, *Bacillus anthracis*, Bcbva shares two virulence plasmids with it, PX01 and PX02 (Klee et al., 2010). Both bacteria show vegetative and endospore stages in their life cycle, with clear evidence for replication inside mammal carcasses. Yet, it is still unclear how similar their ecologies are. Indeed, even the apparent restriction of Bcbva to forested areas is uncertain—a recent seroprevalence study detected anti-Bcbva antibodies in the human population living around TNP but failed at linking seropositivity to forest use or contact to wildlife (Dupke et al., 2020).

Fly iDNA has clear potential to contribute to understanding the restriction of Bcbva to forest ecosystems and aspects of this pathogen's ecology. For classic anthrax, flies have been implicated as a potential mechanical vector and the high proportion of flies carrying viable Bcbva (~5% contained Bcbva), particularly in flies associated with monkey groups, supports the notion that flies play a role in spreading Bcbva as well (Fasanella et al., 2010; Gogarten et al., 2019; Hoffmann et al., 2017). Fly iDNA also contributed to understanding the ecology of the pathogen by expanding the putative host range of Bcbva, through detection of the bacterium and the DNA of a diversity of mammal species in specific flies (Hoffmann et al., 2017). Fly iDNA also helped to understand the distribution of the pathogen at

various spatial scales, from the prevalence of Bcbva across TNP, to an expansion of its known geographic distribution to include Liberia (Hoffmann et al., 2017).

Here, we use fly iDNA to explore the distribution of Bcbva along a gradient from the TNP forest to the surrounding villages. Using the same iDNA, we also aimed to reconstruct fly and mammal communities along this gradient with the aim of investigating potential links with Bcbva detection.

2 | MATERIALS AND METHODS

2.1 | Study site and sample collections

Flies were trapped on a gradient spanning from pristine forest within TNP to the surrounding villages on alternate days from the end of July to the beginning of September 2021, during the rainy season. A total of nine trapping areas were located along three parallel transects, covering three different habitats. Specifically, three trapping areas were selected in villages' periphery, with traps located ~100 m away from the last house of the three villages (hereafter referred to as village habitat; Figure 1). Three areas were selected within the transition area between cultivated land and TNP, where degraded forest encountered cocoa fields or manioc/banana fields (hereafter referred to as the edge habitat; Figure 1). Lastly, three areas were selected well into the pristine forest within TNP, in a mix of swampy and dry primary forest (hereafter referred to as forest habitat; Figure 1).

Fly trapping was carried out for 20–30 min, depending on the trapping success, with the aim of collecting 40 flies per sampling event. Flies were trapped on the ground using custom-made nets (for a detailed description, see Calvignac-Spencer et al., 2013; Hoffmann et al., 2018). Internal organs of cow and pig were purchased from the local market and used as bait and a piece of net minimized direct contact between flies and the bait. Previous analyses in TNP of the fly species attracted with a trapping approach using a piece of a decaying animal or a commercially available bait based on animal proteins that mimic a decaying carcass, described a diversity of dipterans from three families (Calliphoridae, Sarcophagidae, and Muscidae; Calvignac-Spencer et al., 2013; Hoffmann et al., 2017, Gogarten et al., 2019). Flies were euthanized with ether and stored on silica at room temperature in 15 mL falcon tubes containing up to 10–12 flies from a single trap ($N = 2350$). We selected 100 fly pools (5 random flies were selected from a single trap/pool) for this study; 25 pools were from the forest, 50 pools from the edge, and 25 fly pools from the village. We elected to concentrate sampling on the edge, as this was the habitat that was the least well characterized in terms of Bcbva prevalence and the mammal community present.

2.2 | DNA extraction

The DNA was extracted from these fly pools using the GeneMatrix Stool DNA purification kit (Roboklon, Berlin, Germany). In brief, five flies were transferred to a bead tube and sliced into smaller pieces with autoclaved scissors and then homogenized using the tissue lyser II (Qiagen). After centrifugation, supernatants were subsequently processed according to the manufacturer's protocol. DNA concentrations were then measured using the Qubit 3 fluorometer and the high sensitivity dsDNA assay kit (Invitrogen by Thermo Fisher Scientific).

2.3 | Bcbva analyses

2.3.1 | Detection using real time PCR

To determine Bcbva presence, fly pool extracts were screened with three quantitative PCR assays targeting three different gene markers (Hoffmann et al., 2017). All extracts were tested for the first gene marker, *pag* (protective antigen gene, located on the pXO1 plasmid; Ellerbrok et al., 2002). Samples that tested positive for the *pag* gene were subsequently screened for the presence of the second gene marker *capB* (gene for capsule synthesis; located on the

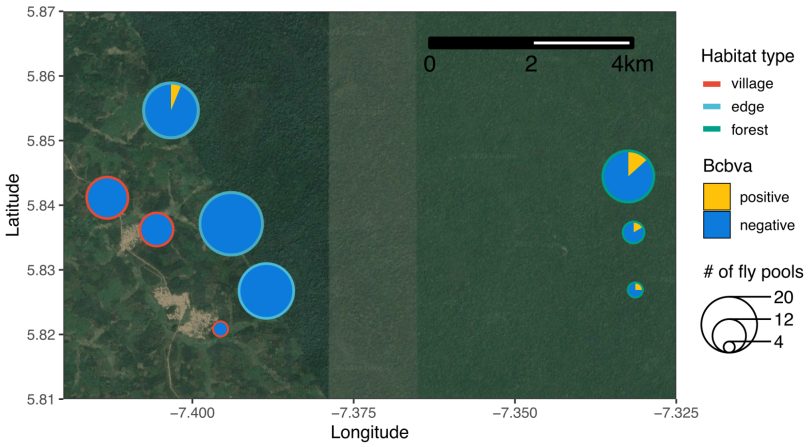

FIGURE 1 Locations of fly trapping along the park boundary of Taï National Park. The area of the circles corresponds to the number of fly pools examined at each location, while the pie chart indicates the number of fly pools positive or negative for Bcbva. The color of the edges of the pie charts corresponds to the habitat type where sampling occurred.

pXO2 plasmid) and the third marker, *Island IV* (chromosomal marker) that is specific for Bcbva, allowing to differentiate Bcbva and *Bacillus anthracis* (Klee et al., 2010). Like Hoffmann et al. (2017), we elected to use three quantitative PCR assays, rather than just a single assay for *Island IV*, to reduce false positives and increase certainty that we were detecting Bcbva and not Bcbva (or another hypothetical and yet unknown *Bacillus cereus* harboring *Island IV*) lacking virulence plasmids (a possibility that cannot be excluded but that would likely affect the bacterium's pathogenic potential). All assays were carried out with two PCR replicates per extract and only considered positive if both replicates were positive. Positive results were further confirmed by culture and whole genome sequencing (see below).

PCRs were carried out in a total volume of 25 μL, consisting of 2.5 mM dUTPs, 50 mM $MgCl_2$, 10 μM of each primer, 10 μM of the analogous probe, 1.25 U Platinum® Taq Polymerase (Invitrogen), 2.5 μL 10× Rxn buffer (Invitrogen), and molecular grade water. The whole reaction was seeded with 200 ng DNA ($N = 16$) or 5 μL of DNA extracts when DNA concentration was below 40 ng/μl ($N = 84$). The following cycling conditions were used: 95°C for 10 min, followed by 45 cycles at 95°C for 15 s and at 60°C for 34 s. Quantitative PCR was carried out using the AriaMX Real-Time PCR system (Agilent Technologies) and analyzed using the Agilent AriaMx software system.

2.3.2 | Culture from qPCR positive isolates

Bacterial culture was conducted via dilution streaking or spread plating of 10 μL fly pool-PBS supernatant onto R & F® *Bacillus cereus/Bacillus thuringiensis* Chromogenic Plating Medium (R & F Products) and incubation at 37°C overnight. Bcbva suspicious, phospholipase-C deficient (as indicated by white colony growth) single colonies were picked.

2.3.3 | Whole genome sequencing of Bcbva isolates

DNA extraction

DNA extraction was performed using the DNeasy Blood and Tissue kit, following the manufacturer's protocol for Gram-positive bacteria (Qiagen). Briefly, Bcbva suspicious colonies were individually picked for DNA extraction. Each colony was transferred into an extraction tube containing 162 μL of lysis buffer. Then 18 μL of lysozyme was added to each extraction tube and the colony was resuspended. This bacterial suspension was then extracted following the manufacturer's protocol for Gram-positive bacteria.

Bacterial DNA concentrations were measured with a Qubit 3 fluorometer and the high sensitivity dsDNA assay kit (Invitrogen by Thermo Fisher Scientific).

Library preparation and sequencing

Bcbva libraries were prepared for whole genome sequencing using the Nextera XT DNA library prep kit (Illumina). Following the manufacturer's protocol, 1 ng of genomic DNA was tagmented, indexed with the Nextera XT index primers and PCR amplified. The cycling conditions were as follows: 72°C for 3 min, 95°C for 30 s, 12 cycles of 95°C for 10 s, 55°C for 30 s, 72°C for 30 s, and a final step of elongation of 72°C for 5 min. Amplified products were then cleaned up using MagSi NGSPREP Plus beads (Steinbrenner Laborsysteme GmbH) to purify the library DNA and short library fragments were removed. Libraries were then quantified using a Qubit, normalized and pooled for sequencing on an Illumina NextSeq 2000 with P2 reagents and 2 × 150 cycles.

2.3.4 | Mammal and fly analyses

DNA extracts were also used to reconstruct the mammals and fly species diversity in and across the TNP using a metabarcoding approach. For mammal metabarcoding, a three-step PCR assay was used to reduce the amplification biases as described in Hoffmann et al. (2017). For the first PCR assay, 130 bp of 16 S mitochondrial DNA was PCR amplified using two universal primers 16 S mam1(5′-CGGTTGGGGTGACCTCGGA-3′), 16 S mam2 (5′-GCT GTT ATC CCT AGG GTA ACT-3′) along with the two sets of blocking primers to reduce the amplification potential laboratory contamination from human and pig, 16 S mam_blk hum 3 (5′-CGGTTGGGGCGACC TCGGAGCAGAACCC-3′) and 16Smam_blkpig (5′-CGGTTGGGGT GACCTCGGAGTACAAAAAAC-3′), respectively. PCRs were carried out with four PCR replicates for all samples and negative template controls. Each PCR was carried out in a total volume of 25 μL, where the reaction was seeded with 200 ng of DNA ($N = 16$) or 5 μL DNA if the DNA concentration < 40 ng/μL ($N = 84$). The reaction mixture included 2.5 mM dNTP (replaced by dUTP), 50 mM $MgCl_2$, 10 μM of each primer, 10 μM of two blocking primers (human and pig), 0.3 U Amperase® uracil N-glycosylase (Invitrogen), 1.25 U Platinum® Taq Polymerase (Invitrogen), 2.5 μL 10× PCR Buffer (Invitrogen), and molecular grade water. Cycling conditions were as follows: 45°C for 7 min, 95°C for 15 min, 42 cycles of 95°C for 30 s, 64°C for 30 s, 72°C for 1 min, followed by elongation at 72°C for 10 min. PCR-amplified products were then visualized on 1.5% agarose gel. All negative template controls were negative on the gel, but were further processed as positive samples. For the latter, products of the expected size were gel-excised under the UV transilluminator and purified using the quick gel purification kit (purelink™; Invitrogen).

Gel-purified products were then submitted to a second PCR. For the PCR (total volume 25 μL), it was seeded with 5 μL of 16 S PCR-product, 2.5 mM dNTP, 4 mM $MgCl_2$, 10 μM of each fusion primer (16Smam primer appended with overhang Illumina specific adapter sequence), 1.25 U Platinum® Taq Polymerase (Invitrogen), and 2.5 10× PCR Buffer (Invitrogen). Cycling conditions were as follows: 95°C for 5 min, 15 cycles of 95°C for 30 s, 64°C for 30 s, 72°C for 1 min, and a final elongation of 72°C for 10 min. These amplified products were cleaned using paramagnetic beads with 1.8 μL × MagSi NGSPREP Plus beads to 1 μL amplicon ratio and eluted in 17.5 μL TET buffer.

In parallel, we also proceeded with fly species metabarcoding, adapting the protocol of Gogarten et al. (2022). Briefly, a 180 bp mitochondrial fragment of cytochrome oxidase C subunit 1 was directly PCR amplified using the ANML primers (Jusino et al., 2019) in fusion with Illumina adapters (total length ~ 250 bp). The primers were LC1490_adapter (5′-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGGTCAACAAATCATAAAGATATTGG-3′) and CO1-CFMRa_Adapter (5′-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGWACTAATCAATTTCCAAATCC-3′). Reaction mixture included 2 μL of DNA extract, 2.5 mM dNTP, 4 mM $MgCl_2$, 10 μM of each fusion primer, 1.25 U Platinum® Taq Polymerase (Invitrogen), and 2.5 μL of 10× PCR Buffer (Invitrogen) in a total volume of 15 μL. Cycling conditions were: 95°C for 5 min, 5 cycles of 94°C for 60 s, 45°C for 90 s, 72°C for 90 s followed by 35 cycles of 94°C for 60 s, 50°C for 90 s, 72°C for 60 s, and a final elongation of 72°C for 7 min. The PCR amplicons were then cleaned up using paramagnetic beads (MagSi NGSPREP Plus beads).

Amplicons from both the fly and mammal metabarcoding from each sample were then pooled and dual-indexed using a Nextera XT Index kit. Briefly, each of the 12.5 μL PCR mixtures contained 1.25 μL of bead-purified PCR amplicons, 6.25 μL 2× KAPA HiFi HotStart ReadyMix (peqLab), 1.25 μL of each Nextera XT index primer (Illumina) and 2.5 μL of nuclease-free water. Cycling conditions were 95°C for 3 min, 8 cycles of 95°C for 30 s, 55°C for 30 s, 72°C for 30 s, and a final elongation step of 72°C for 5 min. The amplicons were further purified using 1.8 μL × MagSi NGSPREP Plus beads to 1 μL amplicon ratio and eluted in a 17.5 μL TET buffer. The indexed amplicon libraries were then quantified using a Qubit 3 and the high sensitivity dsDNA assay kit (Invitrogen) and sequenced on an Illumina Nextseq 500 with a mid-output kit v.2 and 2 × 150 cycles.

2.4 | In silico analyses

2.4.1 | Bcbva genomic analysis

Previously published whole genome sequencing data from Bcbva isolates collected in TNP were downloaded from the Sequence Read Archive (SRA) and converted to fastq files using the sra-toolkit (v. 2.22.0; Leinonen et al., 2011). Fastp (v.0.12.4) was used for read adapter trimming and quality assessment (Chen et al., 2018). Genomic variants were called via the snippy tool (v. 4.6.0; https://github.com/tseemann/snippy), comparing genomes to the Bcbva CI typestrain (GCA_000143605.1) and a variant-site multi fasta alignment was generated using SNP-sites (v.2.2.3; Page et al., 2016). The DNA evolution model was selected via modeltest-ng (v. 0.1.7) and the tree was inferred in a maximum likelihood framework with the help of raxml-ng (v.1.1) starting from 20 random and 20 parsimony trees (Darriba et al., 2020; Kozlov et al., 2019). Branch support values of the best ML tree were evaluated by performing 1000 bootstraps and calculating Transfer Bootstrap Expectation values (Lemoine et al., 2018).

2.4.2 | Mammal metabarcoding

We joined paired-end raw reads using the *illuminapairedend* command of the OBITools package (v1.2.13), setting the minimum alignment score to 40 and removing any non-overlapping reads (Boyer et al., 2016). Primer sequences were removed using the *ngsfilter* command in OBITools and then quality-trimmed with Trimmomatic (v0.36), using a minimum quality score of 30 over a sliding window of four bases, as well as a leading and trailing minimum quality score of 30 and a minimum surviving read length of 80 bp. We then de-replicated the surviving reads using the OBITools *obiuniq* command. We built a reference database using the OBITools *ecoPCR* (v0.2) command to run an in silico PCR on all mammal sequences in GenBank, allowing three mismatches between primers and reference sequences and a synthetic product length between 50 and 800 bp. We then used this database to assign a taxonomy to surviving reads from our experiment using the OBITools *ecotag* command, with a minimum identity level of 0.97. Downstream analyses were based on sequences that were assignable to a mammal species.

Negative controls contained only a single read that could be assigned to the species level, but to conservatively avoid false positives, we only considered a species present when it was detected with at least 10 reads and the species represented at least 0.1% of the total assignable reads for the pool. In addition, the pipeline dropped reads assigned to domestic genera and frequent laboratory contaminants (i.e., those assigned to the genera *Bos*, *Canis*, *Sus*, *Equus*, and *Homo*), as well as reads that were assigned to the genus *Macaca*, for which a closer investigation with blast revealed that these represented a hit to a bacterial artificial chromosome and not the targeted 16 S barcoding region.

2.4.3 | Fly metabarcoding

Primers were removed with cutadapt v2.1 (Martin, 2011) and assigned to amplicon sequence variants (ASVs) using the DADA2 pipeline (Callahan et al., 2016). Sequences trimmed for quality using the filtering parameters (maxN = 0, truncQ = 2, rm.phix = TRUE and maxEE = 2, minLen = 20, trimRight = 1) with the *filterAndTrim* function and then denoised, merged, and chimeras removed with the DADA2 pipeline implemented in R (Callahan et al., 2016). Reads were then assigned to taxa using the eukaryote CO1 reference set v4.0 with the RDP classifier (Porter & Hajibabaei, 2018; Wang et al., 2007). We followed the suggestion of Porter and Hajibabaei (2018) and used bootstrap support cutoff value of 0.6 as this was shown to produce at least 99% correct assignments with barcodes of this length (Porter & Hajibabaei, 2018). To focus on flies, we then considered only ASVs that were assigned to the family Diptera at this threshold. Negative controls contained only two reads that were assigned to the family Diptera, but to conservatively avoid false positives, we considered a ASV present only when it represented at least 5% of the Dipteran reads for a given sample and at least 10 reads. This higher threshold compared to the mammal detections was selected as the fly pools

were known to contain five flies, rather than the trace amounts of mammalian DNA found in and on these flies.

Surviving ASVs were then assigned to molecular operational taxonomic units (MOTUs) by first aligning ASVs with Geneious Prime (v2021.2.2) and creating a phylogeny of the COI barcode using IQTREE (V2.2.0.3), testing for the best fitting model with ModelFinder (N = 55 sequences). Bootstrap support was estimated using the ultra-fast bootstrap procedure with 1000 pseudo-replicates. Subsequent analyses focused on the best fitting model according to BIC (TIM2+F+G4). This tree was then used as the input for a general-mixed Yule/coalescent (GMYC) analysis implemented in the R package splits (Fujisawa & Barraclough, 2013), with the aim of delineating putative species (Pons et al., 2006). This approach required the tree be rooted and ultrametric, which was achieved using the *chronos* function in the R package ape (Paradis & Schliep, 2019) and *midpoint.root* function in the R package phytools (Revell, 2012). The GMYC approach resulted in the delineation of 34 fly MOTUs, which formed the basis of subsequent analyses.

2.4.4 | Statistical analysis and visualizations

We used a Fisher's exact to compare the Bcbva positivity in the three habitat types. To test for a difference in the number of mammal or fly species detected per fly pool in the three habitat types, we used a Generalized Linear Model with a Poisson error structure and log link function (McCullagh & Nelder, 1989). We fit a model including habitat as a fixed effect and fit the model in R using the function glm. The sample for this model comprised 100 fly pools. To establish the significance of the full model, we used a likelihood ratio test (Dobson & Barnett, 2002), comparing its deviance with that of a null model comprising only the intercept.

A satellite image of the sampling region was downloaded from the google api and pie charts were plotted on this image using the *ggmap*, *scatterpie*, and *ggplot2* R packages (Kahle & Wickham, 2013; Wickham, 2016). We constructed species accumulation curves of mammal and fly species diversity using the BiodiversityR, R package to calculate the expected mean species richness for different sample sizes (Kindt & Coe, 2005). We constructed Venn diagrams of species overlap between habitat types using the ggvenn R package (Yan, 2021). To account for differences in the number of fly pools tested in the different habitat types, for the Venn diagrams we downsampled the data to the minimum number of fly pools in a habitat type (N = 25). The maximum likelihood phylogeny was plotted with the help of the ggtree package (Yu, 2020).

All statistical analyses were performed in R v4.2.1 (R Core Team, 2021).

3 | RESULTS

We detected Bcbva in 5 of the 100 fly pools tested. Bcbva positivity of fly pools varied significantly across the three habitat types (forest = 4/25, edge = 1/50, and village = 0/25: Fisher's exact test; P = 0.026; Figure 1). It was possible to culture Bcbva from all five fly pools, confirming their positivity. We generated whole genomes from each of these pools with a chromosomal depth of coverage of at least 114 X.

We generated a total of 2,530,239 paired reads from the pooled 16S mammal metabarcoding and COI fly metabarcoding experiment. Of these paired reads, 1,593,484 survived the obitools processing pipeline for mammals and were assigned a taxonomic rank ($\bar{x}_{\text{per fly pool}} \pm \text{SD} = 15{,}934 \pm 9242$ paired reads). The number of mammalian species per fly pool varied by habitat type ($\chi^2 = 62.2$, df = 2, $p < 0.001$; Figure 2a), with more mammal species detected per fly pool in the forest than either the edge or village ($z = 6.47$, $p < 0.001$); similarly, there were more mammal species in the edge than the village ($z = 3.85$, $p < 0.001$). Species accumulation curves revealed that mammal diversity in the village plateaued well below the species diversity observed in the forest and edge (Figure 2b). In both the edge and the forest, the species accumulation curves had not yet plateaued, suggesting more species are yet to be described in these habitats. The initial rate of accumulation of novel species detections was higher in the forest than the edge habitat. There was considerable overlap in the species detected in the edge and village habitats, while none of the mammal species detected in the village were unique to the village. We detected five species in all habitat types (*Cercopithecus campbelli*, Campbell's monkey; *Cercopithecus diana*, Diana monkey; *Cricetomys* sp., giant pouched rat; *Piliocolobus badius*, western red colobus; *Procolobus verus*, olive colobus) and a number of species were detected in both the edge and the forest, but not the village (*Atherurus africanus*, African brush-tailed porcupine; *Crossarchus* sp.; *Crossarchus obscurus*, common kusimanse; *Nandinia binotata*, African palm civet; *Philantomba maxwellii*, Maxwell's duiker; *Potamochoerus porcus*, red river hog), while *Protoxerus stangeri*, the forest giant squirrel, was the only species detected in both the forest and village, but not the edge. Prior to downsampling, there were eight species detected only in the forest (*Anomalurus beecrofti*, Beecroft's flying squirrel; *Anomalurus derbianus*, Lord Derby's scaly-tailed squirrel; *Cephalophus jentinki*, Jentink's duiker; *Cercopithecus nictitans*, greater spot-nosed monkey; Crossarchus sp.; *Dendrohyrax dorsalis*, western tree hyrax; *Hybomys trivirgatus*, Temminck's striped mouse; *Hyemoschus aquaticus*, water chevrotain), while nine species were only detected in the edge (*Cephalophus* sp.; *Cephalophus dorsalis*, bay duiker; *Cercopithecus petaurista*, lesser spot-nosed monkey; *Colobus polykomos*, king colobus; *Crocidura grandiceps*, long-headed shrew; *Eidolon helvum*, straw-colored fruit bat; *Grammomys* sp.; *Lemniscomys striatus*, Typical striped grass mouse; *Mastomys natalensis*, Natal multimammate mouse; *Mops condylurus*, Angolan free-tailed bat; Figure 2c).

All of the nine mammal species detected in Bcbva-positive fly pools (African brush-tailed porcupine, Campbell's monkey, Diana monkey, forest giant squirrel, giant pouched rat, greater spot-nosed monkey, Maxwell's duiker, olive colobus, western red colobus) were also detected in Bcbva-negative fly pools. Most of these species were not specific to the edge or forest, with the exception of greater

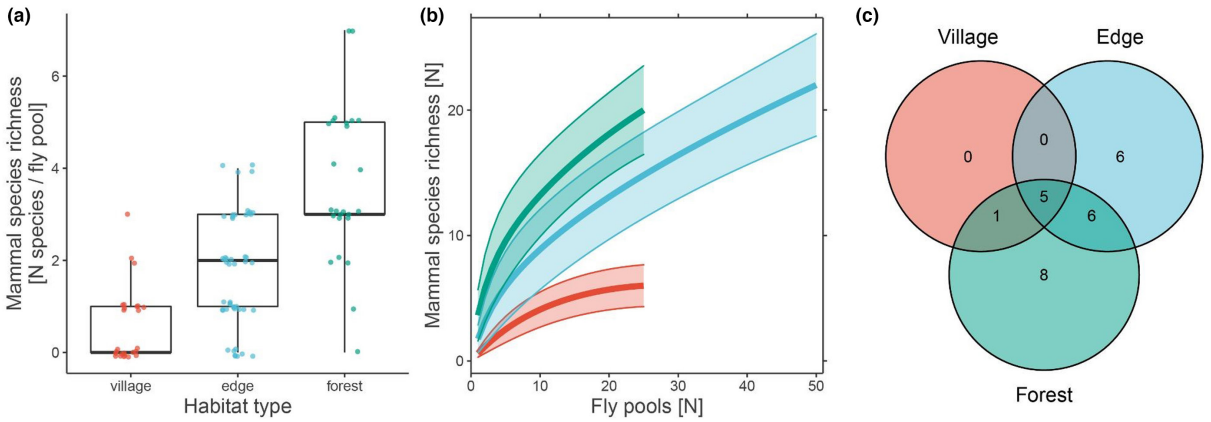

FIGURE 2 (a) A box and whisker plot of the number of mammal species detections per fly pool in the different habitat types. The horizontal line in the whisker plots represents the mean, while the lower and upper bounds of the boxes indicate the first and third quartiles, respectively. The upper and lower whiskers extend to the largest and lowest values no more than 1.5 times the interquartile range away from the hinge. Raw data are plotted as points, with the color corresponding to the habitat type. (b) Mammal species accumulation curves for the three habitat types, with the colors indicating the different habitats as in a and c, with the shade areas indicating the 95% confidence interval. (c) Venn diagrams showing the overlap of the mammal species detected in each of the different habitat types. Data were downsampled for the Venn diagram so that the mammal detections are from the same number of fly pools ($N = 25$) in each habitat type.

spot-nosed monkey and Maxwell's duiker, that were found only in the forest, while the forest giant squirrel was detected in the forest and village. Decades of carcass monitoring in TNP have detected Bcbva in carcasses of *Cephalophus* sp., West African chimpanzees (*Pan troglodytes verus*), Hystricidae, Herpestidae, king colobus, western red colobus, Diana monkeys, lesser spot-nosed monkeys, Campbell's monkeys, and sooty mangabeys (*Cercocebus atys*), which suggest considerable overlap in carcass- and fly-based association of Bcbva and mammal species. Here, the forest giant squirrel, greater spot-nosed monkey, Maxwell's duiker, and giant pouched rat represent potential hosts detected for the first time in Bcbva-positive flies, though are in complete accordance with previous genus-level analyses of mammal detections in Bcbva-positive flies from the forest interior of TNP.

We generated a total of 488,546 paired fly COI metabarcoding reads, of which 445,420 could be assigned a taxonomy with the RDP classifier ($\bar{x}_{\text{per fly pool}} \pm \text{SD} = 4454.18 \pm 2920$ paired reads). Of these, 442,034 were assigned to one of the 34 Dipteran MOTUs ($\bar{x}_{\text{per fly pool}} = 4420 \pm 2915$ paired reads) and 432,065 met the minimum proportion of reads in a sample thresholds ($\bar{x}_{\text{per fly pool}} = 4320 \pm 2869$ paired reads). Fly MOTU detection per fly pool varied significantly by habitat type ($\chi^2 = 7.18$, df = 2, $P = 0.028$; Figure 3a), with more fly MOTUs detected per fly pool in the edge than the village ($z = 2.57$, $P = 0.01$), with no significant difference in the fly MOTUs detected per fly pool in the forest or village. The fly species accumulation curve plateaued in the village habitat suggesting much of the fly MOTU diversity had been described, but in both the edge and forest, the curves were still increasing at the maximum sample size, suggesting that further sampling would continue to reveal novel fly MOTUs in these habitats (Figure 3b). There were four generalist MOTUs that were detected in all habitats; these could be assigned to the families Calliphoridae and Sarcophagidae by the RDP classifier. In addition, there were many novel MOTUs detected only in one habitat (Figure 3b); the 10 MOTUs detected only in the edge belonged to the family Calliphoridae, while the 3 MOTUs detected only in the forest belonged to the family Calliphoridae, and the one MOTU detected only in the village belonged to the family Muscidae (Figure 3c). All six of the fly MOTUs detected in Bcbva-positive fly pools were also detected in Bcbva-negative fly pools and could be assigned to the families Calliphoridae and Sarcophagidae.

The five Bcbva genomes generated here span a considerable portion of known Bcbva diversity in this ecosystem (Figure 4). The one Bcbva detection in a fly captured at the edge revealed a genome identical to one of those sampled in the forest.

4 | DISCUSSION

We used fly iDNA to explore the distribution of Bcbva along a gradient from villages to the forest. At the same time, iDNA generated data about the mammal and fly communities in these habitats, providing insights into the ecology of this pathogen. For both Bcbva and mammal species DNA, the detection rate varied across habitat types. The highest Bcbva detection rate was in the forest, where we also saw the highest mammal species richness and highest mammal detection rate in flies. We did not detect Bcbva in the village habitat, which was also the most mammal and fly species-poor environment. The Bcbva that we detected with iDNA spanned much of the known genomic diversity of this pathogen generated with decades of carcass monitoring (Hoffmann et al., 2017). Similarly, the mammal hosts detected in Bcbva-positive fly pools showed considerable overlap with the mammal species' whose carcasses contained Bcbva

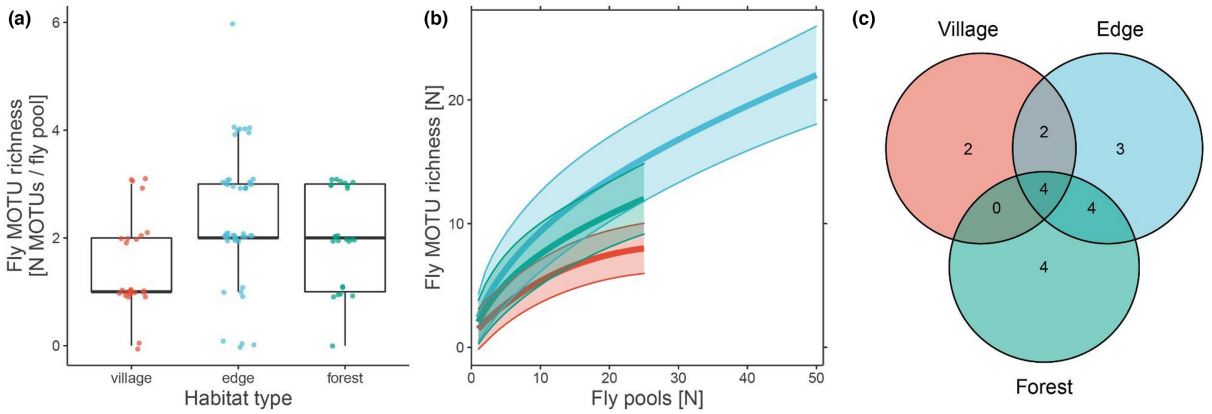

FIGURE 3 (a) A box and whisker plot of the number of fly MOTU detections per fly pool in the different habitat types. The horizontal line in the whisker plots represents the mean, while the lower and upper bounds of the boxes indicate the first and third quartiles, respectively. The upper and lower whiskers extend to the largest and lowest values no more than 1.5 times the interquartile range away from the hinge. Raw data are plotted as points, with the color corresponding to the habitat type. (b) Fly MOTU accumulation curves for the three habitat types, with the colors indicating the different habitats as in a and c, with the shade areas indicating the 95% confidence interval. (c) Venn diagrams showing the overlap of the fly MOTUs detected in each of the different habitat types. Data were downsampled for the Venn diagram so that the fly MOTU detections are from the same number of fly pools ($N = 25$) in each habitat type.

(Hoffmann et al., 2017). Our results highlight the power of iDNA for biomonitoring and pathogen surveillance of emerging and newly described pathogens.

The high Bcbva detection rate in forest flies described here is not unprecedented (4/25 fly pools); indeed previous work in TNP on individual flies, found high Bcbva detection rates in flies forming a long-term association with a group of sooty mangabeys in TNP (~7% of individual flies positive for Bcbva; Gogarten et al., 2019). A broader survey of flies across TNP forest also suggested a high positivity rate in the research area where hunting is rare (~5% of individual flies; Hoffmann et al., 2017). Interestingly, Bcbva positivity varied across the forest habitat and it was more likely to detect Bcbva in flies within the research area than adjoining forest where mammal diversity and biomass was lower because of hunting. In these TNP fly iDNA studies, Bcbva positivity was also more likely in flies that contained more mammal DNA, perhaps indicating recent contact with an animal carcass (Hoffmann et al., 2017). Our findings echo these results, in that Bcbva detection was highest in the forest habitat, where mammal detection rates in fly pools and overall mammal species diversity were highest. We hypothesize that a larger number of mammal species and a higher mammal biomass support more Bcbva mortality in an area, which, in turn, results in more flies coming into contact with Bcbva.

Traditional survey techniques have been used in TNP to monitor populations of certain mammal species and found mammal declines that correlate with signs of hunting (Hoppe-Dominik et al., 2011; Köndgen et al., 2008). While it was hypothesized that the forest edge represents a hostile environment for wild mammals due to ease of access to hunters, this was difficult to quantify in part because animals might also modify their behavior to reduce detection in these hunting areas (Benhaiem et al., 2008). Our iDNA results suggest that mammal biodiversity is considerably lower at the forest edge than a few kilometers away in the forest. Some caution is warranted when interpreting iDNA for biodiversity monitoring, in that detecting the DNA of an animal need not mean the animal was alive in that habitat (i.e., DNA could originate from wildlife carcasses hunted in the forest but processed in the village). Indeed, the detection of non-human primate DNA in village flies might be indicative of the regular consumption of these species in the region (Refisch & Koné, 2005). While camera traps, acoustic monitoring, or dung counts can also be useful for detection of certain species, our findings suggest that iDNA represents a useful tool for describing biodiversity in edge and village ecosystems where mammal monitoring with traditional surveys can be more challenging.

To reduce costs and increase our sample sizes, we used a fly pooling strategy. Pooling has drawbacks when studying pathogen ecology, in particular when looking for mammal host–pathogen detection correlations to assign a pathogen to a particular host or vector species, since it is not possible to make sure that host and pathogen DNA stem from a same individual fly (Alfano et al., 2021; Mwakasungula et al., 2022). This problem is even complicated by potential DNA movements in the collection tube before the analyses, which might increase the frequency of spurious codetection events. These limitations can be compensated with larger sample sizes that enable more detections and allow the use of modeling approaches to identify codetections across multiple pools; this was not feasible here as we only had five Bcbva detections. Despite these challenges, the high degree of overlap between iDNA host species estimates and those generated from carcass monitoring suggests this pooling strategy can generate biologically meaningful results in keeping with the extremely broad host range known for this pathogen (Hoffmann et al., 2017).

In contrast to mammal diversity, fly MOTU diversity was higher at the forest edge than the forest or villages. A number of factors

Environmental DNA
Dedicated to the study and use of environmental DNA for basic and applied sciences
Open Access
WILEY

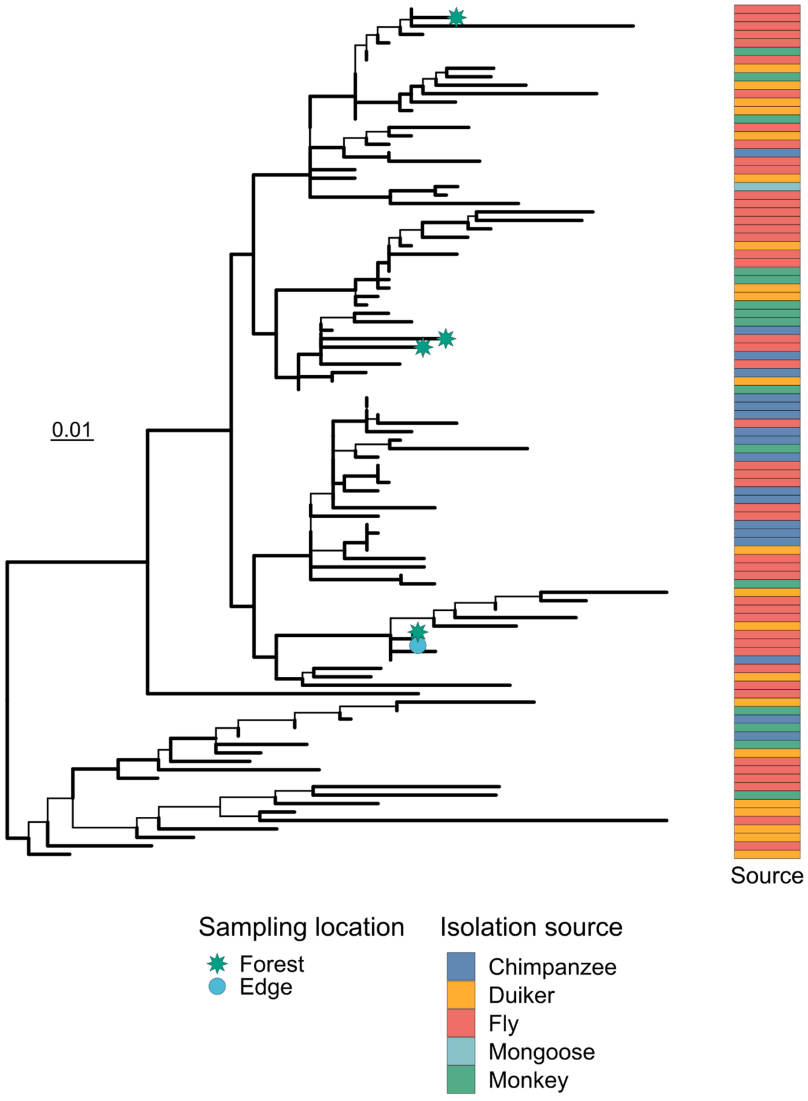

FIGURE 4 Maximum likelihood phylogeny of Bcbva chromosomes collected in TNP. The tree is midpoint rooted and internal branches with transfer bootstrap values below 0.7 are displayed in non-bold. Tip points are displayed for the sequences generated within this study, whereas the points color and shape represent the fly sampling location. The isolation sources of all sequences are illustrated by a colored strip. Substitutions per site are represented by the scale bar.

might be involved in driving this pattern, but the habitat heterogeneity at the edge may support the presence of both more anthropogenic flies, those that thrive in more savannah like ecosystems, as well as forest specialists. This is in keeping with broader studies of invertebrate biodiversity at edges, that suggest even sharp habitat edges as defined by major changes in soils and plants do not serve as hard edges but broad transition zones for many invertebrate taxa (Dangerfield et al., 2003). Indeed, forest edges in a tropical forest in Indonesia showed higher levels of insect diversity than the forest interior (Darsono et al., 2020). Whether particular fly species are more likely to carry Bcbva is unclear, but the variation in fly species community composition might be linked to the variation in Bcbva detection rates across habitats.

Flies have been implicated as vectors for a number of pathogens, including *Bacillus anthracis*, the causative agent of classic anthrax (Blackburn et al., 2010; Greenberg, 1971; Turell & Knudson, 1987). Our finding of a Bcbva-positive fly pool at the forest edge, in areas frequently used by people, may represent a route of exposure to Bcbva. Flies can travel large distances, at the scales examined here; for example, a mark–recapture experiment in TNP showed that flies moved at least 1.3 km in the forest with a monkey group (Gogarten et al., 2019). Similarly, flies can move at least a few 100 m across the forest edge into village areas around Kibale National Park, Uganda suggesting they may serve as mechanical vector between these ecosystems (Jahan et al., 2023). While the potential mobility of flies may impact the scale at which pathogen and mammal iDNA-based monitoring is meaningful, the biological signal extracted from the mammal and fly diversity estimates across these habitats suggests these fly populations are not a homogeneous population mixture at this spatial scale.

We detected an identical Bcbva isolate in a fly captured in the forest and edge. We previously described Bcbva diversity within carcasses and found that isolates differed by a maximum of two chromosomal single nucleotide polymorphisms (Hoffmann et al., 2017),

suggesting these flies were likely exposed at the same carcass or epidemiologically linked carcasses. The lack of Bcbva detection in village areas suggests that Bcbva is not causing mammal mortality in this habitat at the scale seen in neighboring forests. While the seropositivity of people to Bcbva shows that exposure happens in this region, how this exposure happens is unclear and determining the impact of Bcbva on people and their livestock represents an important area of research.

Our results demonstrate the power of iDNA for biomonitoring and pathogen surveillance along gradients of anthropogenic disturbance. Fly iDNA revealed the spatial scope and host range of Bcbva at the forest edge and surrounding villages and supported a strict reliance of this bacterium on the rainforest ecosystem.

AUTHOR CONTRIBUTIONS

MJ, LL, TG, ECH, FHL, SCS, and JFG contributed to the conception or design of the study and MJ, LL, TG, SCS, and JFG prepared the first draft of the manuscript. All authors contributed to the acquisition, analysis, or interpretation of the data, as well as contributed to editing and shaping the final manuscript.

ACKNOWLEDGMENTS

For their support and permission to conduct this study, we would like to thank the Ministère de l'Enseignement Supérieur et de la Recherche Scientifique, the Centre de Recherche en Ecologie, and the Ministère des Eaux et Fôrets in Côte d'Ivoire, as well as the Office Ivoirien des Parcs et Réserves. We also thank the Sequencing Unit of the Robert Koch Institute for their assistance with this project. We are grateful to the Centre Suisse de Recherches Scientifiques en Côte d'Ivoire for their support in organizing and actualizing this project. Work in the Côte d'Ivoire was supported by the EU-Biodiversa BIODIV-AFREID project (LE 1813/17-1: Project number: 428839112). MJ was funded by the Bangabandhu Science and Technology Fellowship Trust, Ministry of Science and Technology, Government of the People's Republic of Bangladesh. JFG was supported by the DFG grant "The ecology and evolution of primate phageomes" (GO 3443/1-1: Project number: 453352748).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Next-generation sequencing reads used for the reconstruction of Bcbva genomes are available through the short read archive (project accession number: PRJNA940060). All additional metabarcoding sequences generated for this study have been uploaded to Zenodo, with the logic that assignment to a particular organism is always uncertain and should not supersede existing sequences linked to a particular host in available sequence databases (doi: 10.5281/zenodo.7688127).

ORCID

Mueena Jahan https://orcid.org/0000-0003-1224-9615
Lorenzo Lagostina https://orcid.org/0000-0002-6661-705X
Tobias Gräßle https://orcid.org/0000-0001-9103-3139
Emmanuel Couacy-Hymann https://orcid.org/0000-0003-2593-6932
Valère K. Kouakou https://orcid.org/0000-0002-5024-1721
Livia V. Patrono https://orcid.org/0000-0003-3978-3607
Kamilla Pléh https://orcid.org/0000-0002-2528-5482
Fabian H. Leendertz https://orcid.org/0000-0002-2169-7375
Sébastien Calvignac-Spencer https://orcid.org/0000-0003-4834-0509
Jan F. Gogarten https://orcid.org/0000-0003-1889-4113

REFERENCES

Alfano, N., Dayaram, A., Axtner, J., Tsangaras, K., Kampmann, M. L., Mohamed, A., Wong, S. T., Gilbert, M. T. P., Wilting, A., & Greenwood, A. D. (2021). Non-invasive surveys of mammalian viruses using environmental DNA. *Methods in Ecology and Evolution*, *12*(10), 1941–1952. https://doi.org/10.1111/2041-210X.13661

Andruszkiewicz, E. A., Starks, H. A., Chavez, F. P., Sassoubre, L. M., Block, B. A., & Boehm, A. B. (2017). Biomonitoring of marine vertebrates in Monterey Bay using eDNA metabarcoding. *PLoS One*, *12*(4), 1–20. https://doi.org/10.1371/journal.pone.0176343

Benhaiem, S., Delon, M., Lourtet, B., Cargnelutti, B., Aulagnier, S., Hewison, A. J. M., Morellet, N., & Verheyden, H. (2008). Hunting increases vigilance levels in roe deer and modifies feeding site selection. *Animal Behaviour*, *76*(3), 611–618. https://doi.org/10.1016/j.anbehav.2008.03.012

Bitome-Essono, P. Y., Ollomo, B., Arnathau, C., Durand, P., Mokoudoum, N. D., Yacka-Mouele, L., Okouga, A. P., Boundenga, L., Mve-Ondo, B., Obame-Nkoghe, J., Mbehang-Nguema, P., Njiokou, F., Makanga, B., Wattier, R., Ayala, D., Ayala, F. J., Renaud, F., Rougeron, V., Bretagnolle, F., … Paupy, C. (2017). Tracking zoonotic pathogens using blood-sucking flies as 'flying syringes. *eLife*, *6*, 1–21. https://doi.org/10.7554/eLife.22069

Blackburn, J. K., Curtis, A., Hadfield, T. L., O'Shea, B., Mitchell, M. A., & Hugh-Jones, M. E. (2010). Confirmation of *bacillus anthracis* from flesh-eating flies collected during a West Texas anthrax season. *Journal of Wildlife Diseases*, *46*(3), 918–922. https://doi.org/10.7589/0090-3558-46.3.918

Bohmann, K., Evans, A., Gilbert, M. T. P., Carvalho, G. R., Creer, S., Knapp, M., Yu, D. W., & de Bruyn, M. (2014). Environmental DNA for wildlife biology and biodiversity monitoring. *Trends in Ecology & Evolution*, *29*(6), 358–367. https://doi.org/10.1016/j.tree.2014.04.003

Boyer, F., Mercier, C., Bonin, A., Le Bras, Y., Taberlet, P., & Coissac, E. (2016). OBITOOLS: A unix-inspired software package for DNA metabarcoding. *Molecular Ecology Resources*, *16*(1), 176–182. https://doi.org/10.1111/1755-0998.12428

Butchart, S. H. M., Walpole, M., Collen, B., Van Strien, A., Scharlemann, J. P. W., Almond, R. E. A., Baillie, J. E. M., Bomhard, B., Brown, C., Bruno, J., Carpenter, K. E., Carr, G. M., Chanson, J., Chenery, A. M., Csirke, J., Davidson, N. C., Dentener, F., Foster, M., Galli, A., … Watson, R. (2010). Global biodiversity: Indicators of recent declines. *Science*, *328*(5982), 1164–1168. https://doi.org/10.1126/science.1187512

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, *13*(7), 581–583. https://doi.org/10.1038/nmeth.3869

Calvignac-Spencer, S., Merkel, K., Kutzner, N., Kühl, H., Boesch, C., Kappeler, P. M., Metzger, S., Schubert, G., & Leendertz, F. H. (2013). Carrion fly-derived DNA as a tool for comprehensive and

cost-effective assessment of mammalian biodiversity. *Molecular Ecology*, *22*(4), 915–924. https://doi.org/10.1111/mec.12183

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, *34*(17), i884–i890. https://doi.org/10.1093/bioinformatics/bty560

Dangerfield, J. M., Pik, A. J., Britton, D., Holmes, A., Gillings, M., Oliver, I., Briscoe, D., & Beattie, A. J. (2003). Patterns of invertebrate biodiversity across a natural edge. *Austral Ecology*, *28*(3), 227–236. https://doi.org/10.1046/j.1442-9993.2003.01240.x

Darsono, R., E, Santoso, S., Sudiana, E., Nasution, E. K., Aprilliana, H., & Chasanah, T. (2020). Insect diversity in various distances to forest edge in small nature reserve: A case study of BantarbolangNature Reserve, Central Java, Indonesia. *Biodiversitas*, *21*(10), 4821–4828.

Darriba, D., Posada, D., Kozlov, A. M., Stamatakis, A., Morel, B., & Flouri, T. (2020). ModelTest-NG: A new and scalable tool for the selection of DNA and protein evolutionary models. *Molecular Biology and Evolution*, *37*(1), 291–294. https://doi.org/10.1093/molbev/msz189

Dobson, A. J., & Barnett, A. G. (2002). *An introduction to generalized linear models* (Second ed.). CRC Press LLC.

Dupke, S., Schubert, G., Beudjé, F., Barduhn, A., Pauly, M., Couacy-Hymann, E., Grunow, R., Akoua-Koffi, C., Leendertz, F., & Klee, S. (2020). Serological evidence for human exposure to *Bacillus cereus* biovar *anthracis* in the villages around Taï National Park, Côte D'ivoire. *PLoS Neglected Tropical Diseases*, *14*(5), e0008292. https://doi.org/10.1371/journal.pntd.0008292.t001

Ellerbrok, H., Nattermann, H., Özel, M., Beutin, L., Appel, B., & Pauli, G. (2002). Rapid and sensitive identification of pathogenic and apathogenic *bacillus anthracis* by real-time PCR. *FEMS Microbiology Letters*, *214*(1), 51–59. https://doi.org/10.1016/S0378-1097(02)00837-6

Fasanella, A., Scasciamacchia, S., Garofolo, G., Giangaspero, A., Tarsitano, E., & Adone, R. (2010). Evaluation of the house fly *Musca domestica* as a mechanical vector for an anthrax. *PLoS One*, *5*(8), 4–8. https://doi.org/10.1371/journal.pone.0012219

Ficetola, G. F., Miaud, C., Pompanon, F., & Taberlet, P. (2008). Species detection using environmental DNA from water samples. *Biology Letters*, *4*(4), 423–425. https://doi.org/10.1098/rsbl.2008.0118

Fujisawa, T., & Barraclough, T. G. (2013). Delimiting species using single-locus data and the generalized mixed yule coalescent approach: A revised method and evaluation on simulated data sets. *Systematic Biology*, *62*(5), 707–724. https://doi.org/10.1093/sysbio/syt033

Gilbert, N. A., Clare, J. D. J., Stenglein, J. L., & Zuckerberg, B. (2021). Abundance estimation of unmarked animals based on camera-trap data. *Conservation Biology*, *35*(1), 88–100. https://doi.org/10.1111/cobi.13517

Gogarten, J. F., Calvignac-Spencer, S., Nunn, C. L., Ulrich, M., Saiepour, N., Nielsen, H. V., Deschner, T., Fichtel, C., Kappeler, P. M., Knauf, S., Müller-Klein, N., Ostner, J., Robbins, M. M., Sangmaneedet, S., Schülke, O., Surbeck, M., Wittig, R. M., Sliwa, A., Strube, C., … Noll, A. (2020). Metabarcoding of eukaryotic parasite communities describes diverse parasite assemblages spanning the primate phylogeny. *Molecular Ecology Resources*, *20*(1), 204–215. https://doi.org/10.1111/1755-0998.13101

Gogarten, J. F., Düx, A., Mubemba, B., Pléh, K., Hoffmann, C., Mielke, A., Müller-Tiburtius, J., Sachse, A., Wittig, R. M., Calvignac-Spencer, S., & Leendertz, F. H. (2019). Tropical rainforest flies carrying pathogens form stable associations with social nonhuman primates. *Molecular Ecology*, *28*(18), 4242–4258. https://doi.org/10.1111/mec.15145

Gogarten, J. F., Jahan, M., Calvignac-Spencer, S., Chapman, C. A., Goldberg, T. L., Leendertz, F. H., & Rothman, J. M. (2022). The cost of living in larger primate groups includes higher fly densities. *EcoHealth*, *19*(2), 290–298. https://doi.org/10.1007/s10393-022-01597-5

Greenberg, B. (1971). *Flies and disease, Vol. 1: Ecology, classification and biotic associations; flies and disease. Volume II. Biology and disease transmission*. Princeton University Press.

Hale, V. L., Dennis, P. M., McBride, D. S., Nolting, J. M., Madden, C., Huey, D., Ehrlich, M., Grieser, J., Winston, J., Lombardi, D., Gibson, S., Saif, L., Killian, M. L., Lantz, K., Tell, R. M., Torchetti, M., Robbe-Austerman, S., Nelson, M. I., Faith, S. A., & Bowman, A. S. (2022). SARS-CoV-2 infection in free-ranging white-tailed deer. *Nature*, *602*(7897), 481–486. https://doi.org/10.1038/s41586-021-04353-x

Hockings, K. J., Mubemba, B., Avanzi, C., Pleh, K., Düx, A., Bersacola, E., Bessa, J., Ramon, M., Metzger, S., Patrono, L. V., Jaffe, J. E., Benjak, A., Bonneaud, C., Busso, P., Couacy-Hymann, E., Gado, M., Gagneux, S., Johnson, R. C., Kodio, M., … Leendertz, F. H. (2021). Leprosy in wild chimpanzees. *Nature*, *598*(7882), 652–656. https://doi.org/10.1038/s41586-021-03968-4

Hoffmann, C., Merkel, K., Sachse, A., Rodríguez, P., Leendertz, F. H., & Calvignac-Spencer, S. (2018). Blow flies as urban wildlife sensors. *Molecular Ecology Resources*, *18*(3), 502–510. https://doi.org/10.1111/1755-0998.12754

Hoffmann, C., Zimmermann, F., Biek, R., Kuehl, H., Nowak, K., Mundry, R., Agbor, A., Angedakin, S., Arandjelovic, M., Blankenburg, A., Brazolla, G., Corogenes, K., Couacy-Hymann, E., Deschner, T., Dieguez, P., Dierks, K., Düx, A., Dupke, S., Eshuis, H., … Leendertz, F. H. (2017). Persistent anthrax as a major driver of wildlife mortality in a tropical rainforest. *Nature*, *548*(7665), 82–85. https://doi.org/10.1038/nature23309

Hoppe-Dominik, B., Kühl, H. S., Radl, G., & Fischer, F. (2011). Long-term monitoring of large rainforest mammals in the biosphere Reserve of Taï National Park, Côte d'Ivoire. *African Journal of Ecology*, *49*(4), 450–458. https://doi.org/10.1111/j.1365-2028.2011.01277.x

Jahan, M., Calvignac-Spencer, S., Chapman, C. A., Kalbitzer, U., Leendertz, F. H., Omeja, P. A., Sarkar, D., Ulrich, M., & Gogarten, J. F. (2023). The movement of pathogen carrying flies at the human–wildlife interface. *EcoHealth*, *19*, 450–457. https://doi.org/10.1007/s10393-022-01621-8

Jusino, M. A., Banik, M. T., Palmer, J. M., Wray, A. K., Xiao, L., Pelton, E., Barber, J. R., Kawahara, A. Y., Gratton, C., Peery, M. Z., & Lindner, D. L. (2019). An improved method for utilizing high-throughput amplicon sequencing to determine the diets of insectivorous animals. *Molecular Ecology Resources*, *19*(1), 176–190. https://doi.org/10.1111/1755-0998.12951

Kahle, D., & Wickham, H. (2013). Ggmap: Spatial visualization with ggplot2. *R Journal*, *5*(1), 144–161. https://doi.org/10.32614/rj-2013-014

Kindt, R., & Coe, R. (2005). Tree diversity analysis. A manual and software for common statistical methods for ecological and biodiversity studies. In *World Agroforestry Centre (ICRAF)*. World Agroforestry Centre (ICRAF).

Klee, S. R., Brzuszkiewicz, E. B., Nattermann, H., Brüggemann, H., Dupke, S., Wollherr, A., Franz, T., Pauli, G., Appel, B., Liebl, W., Couacy-Hymann, E., Boesch, C., Meyer, F. D., Leendertz, F. H., Ellerbrok, H., Gottschalk, G., Grunow, R., & Liesegang, H. (2010). The genome of a *bacillus* isolate causing anthrax in chimpanzees combines chromosomal properties of *B. cereus* with *B. anthracis* virulence plasmids. *PLoS One*, *5*(7), e10986. https://doi.org/10.1371/journal.pone.0010986

Köndgen, S., Kühl, H., N'Goran, P. K., Walsh, P. D., Schenk, S., Ernst, N., Biek, R., Formenty, P., Mätz-Rensing, K., Schweiger, B., Junglen, S., Ellerbrok, H., Nitsche, A., Briese, T., Lipkin, W. I., Pauli, G., Boesch, C., & Leendertz, F. H. (2008). Pandemic human viruses cause decline of endangered great apes. *Current Biology*, *18*(4), 260–264. https://doi.org/10.1016/j.cub.2008.01.012

Kozlov, A. M., Darriba, D., Flouri, T., Morel, B., & Stamatakis, A. (2019). RAxML-NG: A fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*, *35*(21), 4453–4455. https://doi.org/10.1093/bioinformatics/btz305

Kuisma, E., Olson, S. H., Cameron, K. N., Reed, P. E., Karesh, W. B., Ondzie, A. I., Akongo, M. J., Kaba, S. D., Fischer, R. J., Seifert, S.

N., Munõz-Fontela, C., Becker-Ziaja, B., Escudero-Pérez, B., Goma-Nkoua, C., Munster, V. J., & Mombouli, J. V. (2019). Long-term wildlife mortality surveillance in northern Congo: A model for the detection of Ebola virus disease epizootics. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, *374*(1782), 1–8. https://doi.org/10.1098/rstb.2018.0339

Leendertz, F. H., Ellerbrok, H., Boesch, C., Couacy-Hymann, E., Mätz-Rensing, K., Hakenbeck, R., Bergmann, C., Abaza, P., Junglen, S., Moebius, Y., Vigilant, L., Formenty, P., & Pauli, G. (2004). Anthrax kills wild chimpanzees in a tropical rainforest. *Nature*, *430*(6998), 451–452. https://doi.org/10.1038/nature02722

Leinonen, R., Sugawara, H., & Shumway, M. (2011). The sequence read archive. *Nucleic Acids Research*, *39*(SUPPL. 1), 2010–2012. https://doi.org/10.1093/nar/gkq1019

Lemoine, F., Domelevo Entfellner, J. B., Wilkinson, E., Correia, D., Dávila Felipe, M., De Oliveira, T., & Gascuel, O. (2018). Renewing Felsenstein's phylogenetic bootstrap in the era of big data. *Nature*, *556*(7702), 452–456. https://doi.org/10.1038/s41586-018-0043-0

Lynggaard, C., Bertelsen, M. F., Jensen, C. V., Johnson, M. S., Frøslev, T. G., Olsen, M. T., & Bohmann, K. (2022). Airborne environmental DNA for terrestrial vertebrate community monitoring. *Current Biology*, *32*(3), 701–707.e5. https://doi.org/10.1016/j.cub.2021.12.014

Martin, M. (2011). CUTADAPT removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*, 10–12. https://doi.org/10.14806/ej.17.1.200

McCullagh, P. & Nelder, J. A. (1989). *Generalized Linear Models* (2nd ed.). Chapman and Hall. https://doi.org/10.1007/978-1-4899-3242-6

Mombo, I. M., Fritz, M., Becquart, P., Liegeois, F., Elguero, E., Boundenga, L., Mebaley, T. N., Prugnolle, F., Maganga, G. D., & Leroy, E. M. (2020). Detection of ebola virus antibodies in fecal samples of great apes in Gabon. *Viruses*, *12*(12), 1–10. https://doi.org/10.3390/v12121347

Morin, D. J., Kelly, M. J., & Waits, L. P. (2016). Monitoring coyote population dynamics with fecal DNA and spatial capture–recapture. *Journal of Wildlife Management*, *80*(5), 824–836. https://doi.org/10.1002/jwmg.21080

Mwakasungula, S., Rougeron, V., Arnathau, C., Boundenga, L., Miguel, E., Boissière, A., Jiolle, D., Durand, P., Msigwa, A., Mswata, S., Olotu, A., Sterkers, Y., Roche, B., Killeen, G., Cerqueira, F., Bitome-Essono, P. Y., Bretagnolle, F., Masanja, H., Paupy, C., … Prugnolle, F. (2022). Using haematophagous fly blood meals to study the diversity of blood-borne pathogens infecting wild mammals. *Molecular Ecology Resources*, *0–13*, 2915–2927. https://doi.org/10.1111/1755-0998.13670

Page, A. J., Taylor, B., Delaney, A. J., Soares, J., Seemann, T., Keane, J. A., & Harris, S. R. (2016). SNP-sites: Rapid efficient extraction of SNPs from multi-FASTA alignments. *Microbial Genomics*, *2*(4), e000056. https://doi.org/10.1099/mgen.0.000056

Paradis, E., & Schliep, K. (2019). Ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, *35*(3), 526–528. https://doi.org/10.1093/bioinformatics/bty633

Pons, J., Barraclough, T. G., Gomez-Zurita, J., Cardoso, A., Duran, D. P., Hazell, S., Kamoun, S., Sumlin, W. D., & Vogler, A. P. (2006). Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Systematic Biology*, *55*(4), 595–609. https://doi.org/10.1080/10635150600852011

Porter, T. M., & Hajibabaei, M. (2018). Automated high throughput animal CO1 metabarcode classification. *Scientific Reports*, *8*(1), 1–10. https://doi.org/10.1038/s41598-018-22505-4

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Refisch, J., & Koné, I. (2005). Impact of commercial hunting on monkey populations in the Taï region, Côte D'ivoire. *Biotropica*, *37*(1), 136–144. https://doi.org/10.1111/j.1744-7429.2005.03174.x

Revell, L. J. (2012). Phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, *3*(2), 217–223. https://doi.org/10.1111/j.2041-210X.2011.00169.x

Rodgers, T. W., Xu, C. C. Y., Giacalone, J., Kapheim, K. M., Saltonstall, K., Vargas, M., Yu, D. W., Somervuo, P., McMillan, W. O., & Jansen, P. A. (2017). Carrion fly-derived DNA metabarcoding is an effective tool for mammal surveys: Evidence from a known tropical mammal community. *Molecular Ecology Resources*, *17*(6), e133–e145. https://doi.org/10.1111/1755-0998.12701

Santiago, M. L., Rodenburg, C. M., Kamenya, S., Bibollet-Ruche, F., Gao, F., Bailes, E., Meleth, S., Soong, S. J., Kilby, J. M., Moldoveanu, Z., Fahey, B., Muller, M. N., Ayouba, A., Nerrienet, E., McClure, H. M., Heeney, J. L., Pusey, A. E., Collins, D. A., Boesch, C., … Hahn, B. H. (2002). SIVcpz in wild chimpanzees. *Science*, *295*(5554), 465. https://doi.org/10.1126/science.295.5554.465

Schnell, I. B., Thomsen, P. F., Wilkinson, N., Rasmussen, M., Jensen, L. R. D., Willerslev, E., Bertelsen, M. F., & Gilbert, M. T. P. (2012). Screening mammal biodiversity using DNA from leeches. *Current Biology*, *22*(8), R262–R263. https://doi.org/10.1016/j.cub.2012.02.058

Stat, M., Huggett, M. J., Bernasconi, R., Dibattista, J. D., Berry, T. E., Newman, S. J., Harvey, E. S., & Bunce, M. (2017). Ecosystem biomonitoring with eDNA: Metabarcoding across the tree of life in a tropical marine environment. *Scientific Reports*, *7*(1), 1–11. https://doi.org/10.1038/s41598-017-12501-5

Stegen, G., Pasmans, F., Schmidt, B. R., Rouffaer, L. O., Van Praet, S., Schaub, M., Canessa, S., Laudelout, A., Kinet, T., Adriaensen, C., Haesebrouck, F., Bert, W., Bossuyt, F., & Martel, A. (2017). Drivers of salamander extirpation mediated by *Batrachochytrium salamandrivorans*. *Nature*, *544*(7650), 353–356. https://doi.org/10.1038/nature22059

Turell, M. J., & Knudson, G. B. (1987). Mechanical transmission of *bacillus anthracis* by stable flies (*Stomoxys calcitrans*) and mosquitoes (*Aedes aegypti* and *Aedes taeniorhynchus*). *Infection and Immunity*, *55*(8), 1859–1861. https://doi.org/10.1128/iai.55.8.1859-1861.1987

Wang, Q., Garrity, G. M., Tiedje, J. M., & Cole, J. R. (2007). Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology*, *73*(16), 5261–5267. https://doi.org/10.1128/AEM.00062-07

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis* (2nd ed.). Springer. https://doi.org/10.1007/978-0-387-98141-3

Yan, L. (2021). ggvenn: Draw Venn Diagram by "ggplot2". R Package Version 0.1.9. https://CRAN.R-project.org/package=ggvenn

Yu, G. (2020). Using ggtree to visualize data on tree-like structures. *Current Protocols in Bioinformatics*, *69*(1), 1–18. https://doi.org/10.1002/cpbi.96

How to cite this article: Jahan, M., Lagostina, L., Gräßle, T., Couacy-Hymann, E., Kouadio, L., Kouakou, V. K., Krou, H. A., Mossoun, A. M., Patrono, L. V., Pléh, K., Steiner, J. A., Yves, N., Leendertz, F. H., Calvignac-Spencer, S., & Gogarten, J. F. (2023). Fly iDNA suggests strict reliance of the causative agent of sylvatic anthrax on rainforest ecosystems. *Environmental DNA*, *00*, 1–12. https://doi.org/10.1002/edn3.401

Linking statement 1

Metabarcoding of invertebrate derived DNA, particularly Fly iDNA has demonstrated growing promise as a rapid and effective tool for assessing wild mammal and vertebrate diversity in tropical and terrestrial ecosystems (Rodgers et al. 2017; Calvignac-Spencer et al. 2013b). In addition, fly iDNA has exhibited potential as a tool for detecting and monitoring emerging infectious pathogens (Hoffmann et al. 2017). By linking both of these sources of information, fly iDNA represents an exciting prospective tool for gaining insights into the ecology and distribution of novel pathogens or emerging infectious diseases. Indeed, in my first chapter, I employed fly iDNA to explore the ecology of a novel pathogen known as *Bacillus cereus* biovar *anthracis* (Bcbva) by studying its geographical distribution and host range along a gradient spanning from pristine forest within Taï National Park, Côte d'Ivoire to the surrounding villages encompassing the forest edge. Interestingly, detection of Bcbva in edge of the forest not only reveals ecological aspect of Bcbva transmission and circulation but also highlights fly-derived iDNA's potential to carry and transmit pathogen from forest to the outward and thus posing a significant risk at the interface between wildlife and humans. This highlights the need for studies exploring the movement of flies at human wildlife interface and an extension of the pathogens these flies might be involved in transmitting.

It has been observed that synanthropic flies that are commonly found near human habitats and livestock have been linked to the transmission of a wide array of pathogens (Greenberg 1973). Recent studies suggest that such fly associations are not unique to humans, with strong evidence that such a persistent association between tropical flies and wild non-human primates, with flies remaining with a primate group for long periods of time as the monkeys move through the forest (Gogarten et al. 2019). This high densities of non-human primate associated flies, which have been observed for many different primate species across the primate phylogeny, could also heighten the risk of disease transmission within primate groups (Gogarten et al. 2022), as these flies have been shown to contain viable Bcbva (Gogarten et al. 2019), while flies in other context have been implicated in the transmission of many pathogens (Junqueira et al. 2017; Förster et al. 2007).

Considering all of these pieces of information in conjunction with the findings of my initial chapter, a pivotal inquiry arises, whether flies maintain a consistent a close association with non-human primates and remain in the rainforest ecosystems, or whether they may move outside of these ecosystems into human use areas, which would mean they have the potential to serve as a vector of wildlife pathogens as they move between habitats. In this next chapter, which was published in *EcoHealth,* I conducted a mark-recapture experiment to investigate the dispersion of flies associated with nonhuman primates across the human use areas

surrounding Kibale National Park, Uganda. Flies from nine groups of nonhuman primates were marked at the forest edge, marking an average of 929 flies per group. Subsequently, recapture effort to attempt to detect these marked flies were performed in human use areas. I also explore the range of potential pathogens that these flies might transmit through metabarcoding. Collectively, results from this chapter illustrate the capacity of flies to serve as vectors between nonhuman primates, livestock and humans at this biodiverse interface.

3.2 Publication II

Jahan, M., Calvignac-Spencer, S., Chapman, C.A. *et al.*

The Movement of Pathogen Carrying Flies at the Human–Wildlife Interface

Published in *EcoHealth* 19, 450-457 (2022)

Received: 07.02.2022

Accepted: 04.10.2022

Published: 11.01.2023

https://doi.org/10.1007/s10393-022-01621-8

EcoHealth 19, 450–457, 2022
https://doi.org/10.1007/s10393-022-01621-8

EcoHealth

Short Communication

The Movement of Pathogen Carrying Flies at the Human-Wildlife Interface

Mueena Jahan,[1,2] Sébastien Calvignac-Spencer,[1,3] Colin A. Chapman,[4,5,6,7] Urs Kalbitzer,[8,9] Fabian H. Leendertz,[1,10] Patrick A. Omeja,[11] Dipto Sarkar,[12] Markus Ulrich,[1,10] and Jan F. Gogarten[1,3,10,13]

[1]*Epidemiology of Highly Pathogenic Organisms, Robert Koch Institute, Berlin, Germany*
[2]*Department of Microbiology and Public Health, Bangabandhu Sheikh Mujibur Rahman Agricultural University, Gazipur, Bangladesh*
[3]*Viral Evolution, Robert Koch Institute Berlin, Berlin, Germany*
[4]*Wilson Center, 1300 Pennsylvania Avenue NW, Washington, DC*
[5]*Center for the Advanced Study of Human Paleobiology, George Washington University, Washington, DC*
[6]*School of Life Sciences, University of KwaZulu-Natal, Pietermaritzburg, South Africa*
[7]*Shaanxi Key Laboratory for Animal Conservation, Northwest University, Xi'an, China*
[8]*Department of Biology, University of Konstanz, Constance, Germany*
[9]*Department for the Ecology of Animal Societies, Max Planck Institute of Animal Behavior, Radolfzell, Germany*
[10]*Helmholtz Institute for One Health, Greifswald, Germany*
[11]*Makerere University Biological Field Station, Fort Portal, Uganda*
[12]*Department of Geography and Environmental Studies, Carleton University, Ottawa, Canada*
[13]*Department of Applied Zoology and Nature Conservation, University of Greifswald, Greifswald, Germany*

Abstract: Flies form high-density associations with human settlements and groups of nonhuman primates and are implicated in transmitting pathogens. We investigate the movement of nonhuman primate-associated flies across landscapes surrounding Kibale National Park, Uganda, using a mark–recapture experiment. Flies were marked in nine nonhuman primate groups at the forest edge ($\bar{x}$ = 929 flies per group), and we then attempted to recapture them in more anthropized areas (50 m, 200 m and 500 m from where marked; 2–21 days after marking). Flies marked in nonhuman primate groups were recaptured in human areas (19/28,615 recaptured). Metabarcoding of the flies in nonhuman primate groups revealed the DNA of multiple eukaryotic primate parasites. Taken together, these results demonstrate the potential of flies to serve as vectors between nonhuman primates, livestock and humans at this biodiverse interface.

Keywords: Disease vector, disease risk, nonhuman primates, disease emergence

Synanthropic flies found in association with human settlements and their livestock have been implicated in the transmission of a large diversity of pathogens (Greenberg 1973). This includes bacteria [e.g., *Chlamydia trachomatis* (Forsey and Darougar 1981)], protozoan parasites [e.g., *Cryptosporidium parvum* (Clavel et al. 2002)], helminths [e.g., *Ascaris lumbricoides* (Adenusi and Adewoga 2013)], as well as viruses [e.g., turkey coronavirus (Calibeo-Hayes et al. 2003)]. Given this potential to serve as disease vectors, higher fly densities are associated with increased disease

Published online: January 11, 2023

Correspondence to: Jan F. Gogarten, e-mail: jan.gogarten@gmail.com

risk (Graczyk et al. 2001; Calibeo-Hayes et al. 2003). The synanthropic flies implicated in increasing disease risk encompass a taxonomically broad and extremely species rich group of Diptera, including a rich diversity from the families Calliphoridae, Sarcophagidae and Muscidae (Greenberg 1973; Stoffolano 2022).

Research suggests that such flies not only form associations with human and livestock populations, but also with wild nonhuman primate groups. For example, fly densities were higher in groups of sooty mangabeys (*Cercocebus atys*), chimpanzees (*Pan troglodytes*), baboons (*Papio anubis*), red colobus (*Piliocolobus tephrosceles*) and black-and-white colobus (*Colobus guereza*), than outside these groups (Gogarten et al. 2019, 2022). A mark–recapture experiment in a group of sooty mangabeys showed that flies can follow a group for up to 13 days, indicative of long-term associations (Gogarten et al. 2019). Particularly in species with a small home range and low daily travel distance, the density of flies in a group increased with larger group sizes (Gogarten et al. 2022). Much like their human-associated counterparts, nonhuman primate-associated flies include a rich diversity of species from the families Calliphoridae, Sarcophagidae and Muscidae (Gogarten et al. 2019, 2022). Collectively, this research suggests that flies form relatively stable associations with a wide range of nonhuman primate species.

These nonhuman primate-associated flies can also carry pathogens and likely increase disease risk. For example, flies associated with a group of sooty mangabeys carried viable *Bacillus cereus* biovar *anthracis*, which causes sylvatic anthrax (Hoffmann et al. 2017; Gogarten et al. 2019). Flies in this group of sooty mangabeys also contained the DNA of *Treponema pallidum pertenue* (Gogarten et al. 2019), which causes yaws disease, which was described in flies from another ecosystems as well (Knauf et al. 2016). Synanthropic flies have been implicated in yaws transmission (Lamboen 1936; Barnard 1952), though it remains unclear to what extent nonhuman primate-associated flies are really involved in the transmission of *Treponema pallidum pertenue* or *Bacillus cereus* biovar *anthracis*. Both *Bacillus cereus* biovar *anthracis* and *Treponema pallidum pertenue* were detected in broad diversity of nonhuman primate-associated fly species; in a subset of 96 flies captured in a group of sooty mangabeys that included 14 putative species, viable *Bacillus cereus* biovar *anthracis* was detected in two fly species, while *Treponema pallidum pertenue* DNA was detected in four other species. Collectively, this suggests that high densities of nonhuman primate-associated flies may pose an increased disease risk by increasing within-group transmission and contamination of substrates that animals come into contact with (Gogarten et al. 2022), though the range of pathogens explored to date remains extremely limited.

The detection of duiker DNA in flies in a sooty mangabey group, as well as the detection of a fly marked in this sooty mangabey group in a chimpanzee group, suggests flies can transfer between animal species and could play a role in between-species transmission as well (Gogarten et al. 2019). This echoes the finding that synanthropic flies can serve as vectors between livestock and humans (Rosef and Kapperud 1983; Khamesipour et al. 2018). Thus, a critical question for nonhuman primate-associated flies at the human wildlife interface is their potential to transmit infectious agents from nonhuman primates (and other wildlife) to livestock and humans. This can be addressed in part by investigating the stability of these associations and the mobility of these flies. A small-scale analysis of the mammalian DNA found in flies in a village near Taï National Park detected the DNA of wildlife species, which could be considered evidence that forest flies enter human habitats (Gogarten et al. 2019). Another possibility though, is that these flies were exposed to the DNA of larger mammal species' through contact with bushmeat in villages, as both duikers and colobines are frequently hunted in this region, while the two rodent species detected are often found in and near human habitats (Refisch and Koné 2005).

Here, we conduct a mark–recapture experiment at the forest edge of Kibale National Park, Uganda to determine whether flies move between wildlife and human populations. We marked flies in nine nonhuman primate groups at the forest edge and tried to recapture these flies in more anthropized areas. To explore the potential disease risk that the movement of nonhuman primate-associated flies into anthropized areas poses, we screened nonhuman primate-associated flies for eukaryotic primate parasites using insect soup metabarcoding (Yu et al. 2012). Kibale National Park consists of a mid-elevation semievergreen forest and contains 13 species of nonhuman primates. We marked 8365 flies in nine groups of four nonhuman primate species that are frequently found at the forest edge close to human settlements (Fig. 1): black-and-white colobus, red colobus, gray-cheeked mangabeys (*Lophocebus albigena*) and red-tailed guenons (*Cercopithecus ascanius*). Nonhuman primate groups were selected opportunistically when they were detected at the forest edge and the number of flies

marked was determined by the fly capture rate once a nonhuman primate group was detected.

Flies were captured using custom-made traps (described in: Hoffmann et al. 2017) placed over a commercial attractant based on animal proteins that mimic a decaying carcass (Unkonventionelle Produkte Feldner, Waldsee, Germany) or a piece of day-old chicken or beef. Flies were marked in large batches in a plastic bag with Glo-Germ powder (Glo Germ Co., Moab, Utah), which can be visualized under UV light. All flies captured during the initial capture event in the nonhuman primate group were marked with powder. Between any two consecutive months of the experiment, we used a different powder color to mark flies in the nonhuman primate groups to avoid incorrect assignment of where flies were marked. To explore fly mobility in anthropized areas, we attempted to recapture flies in the human habitat at a distance of roughly 50 m, 200 m and 500 m from where they were initially marked. Recapture attempts occurred 2, 4, 7, 14 and 21 days after marking, resulting in a total effort of 28,615 flies, with recapture effort dictated by the capture rate at these locations. Flies were checked for Glo Germ powder with a UV light. Kibale National Park is characterized by two rainy and two dry seasons and to explore potential seasonal variation in fly mobility, we compare the monthly rainfall totals assessed immediately adjacent to the study area in months during which recapture occurred and those in which no recapture events occurred (Chapman et al. 2021).

A total of 19 of the 8365 marked flies (0.23%) were recaptured away from the nonhuman primate group in anthropized areas (Fig. 1; Table 1). This included 9 flies at a distance of 50 m (recapture effort = 9681 flies), 8 flies at a distance of 200 m (recapture effort = 9937 flies) and 2 flies at a distance of 500 m (recapture effort = 8997 flies) from where they were marked. Flies marked in groups of three of four species of nonhuman primate examined were recaptured outside of these groups (all except from the gray-

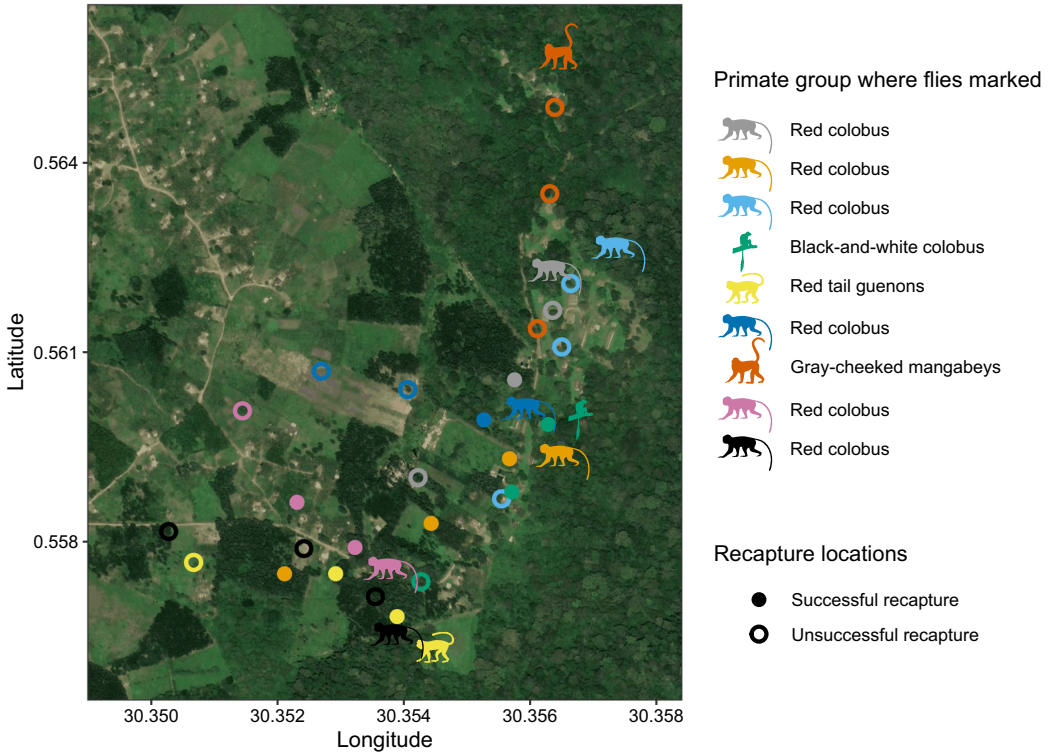

Figure 1. Locations where flies were marked in primate groups and where recapture effort in more anthropized areas occurred. Monkey silhouettes of different colors indicate the group location at the forest edge where flies were marked on a specific day. Circles indicate sites where recapture effort was targeted on subsequent days. Solid filled circles indicate a location where a marked fly was recaptured, while an unfilled circle indicates no flies were recaptured at that location. Colors of circles correspond to recapture effort toward flies marked on a specific date, indicated by the colors of the monkey silhouettes (Color figure online).

Table 1. Number of flies marked and recaptured in six nonhuman primate groups.

Mark date	Primate species	Distance of trap location	N flies marked in primate group	N flies recapture effort	N marked flies recaptured
10/24/20	Red colobus	In group	396		
		50 m		936	0
		200 m		1298	1
		500 m		1324	0
11/20/20	Red colobus	In group	435		
		50 m		982	2
		200 m		1211	3
		500 m		1445	2
1/15/21	Red colobus	In group	580		
		50 m		1517	0
		200 m		1080	0
		500 m		1782	0
2/11/21	Black-and-white colobus	In group	514		
		50 m		1145	2
		200 m		974	1
		500 m		1034	0
3/11/21	Red-tailed guenons	In group	955		
		50 m		2026	1
		200 m		1735	2
		500 m		990	0
4/8/21	Red colobus	In group	1485		
		50 m		1370	3
		200 m		1515	0
		500 m		422	0
5/6/21	Gray-cheeked mangabeys	In group	1270		
		50 m		515	0
		200 m		584	0
		500 m		530	0
6/3/21	Red colobus	In group	1550		
		50 m		840	1
		200 m		1080	1
		500 m		1130	0
7/3/21	Red colobus	In group	1180		
		50 m		350	0
		200 m		460	0
		500 m		340	0
		Total	8365	28,615	19

cheeked mangabey group). Compared to these recapture rates in anthropized areas, a mark–recapture experiment in a mangabey group in Taï National Park, Côte d'Ivoire found a much higher recapture rate (51/1591 = 3.2% of marked flies recaptured; recapture effort = 3164; Gogarten et al. 2019). Collectively these findings suggests that flies preferentially maintain an association with a nonhuman primate social group and maintain these associations, but do occasionally leave these association and move into anthropized areas.

Recapture rates declined with increasing time since flies were marked; 12 flies were recaptured 2 days after marking (recapture effort = 6117 flies) and 7 flies after 4 days (recapture effort = 6469 flies), while no flies were recaptured after 7 days (recapture effort = 4450 flies), after 14 days (recapture effort = 5777 flies) or after 21 days (recapture effort = 5902 flies). The time dependent decay in recapture success may suggest either diffusion of flies, the loss of marking powder through time or relatively short fly survival post-capture (due to the combination of short lifespans and age at capture, capture stress or exposure to the powder), hypotheses that we are not able to differentiate further here. While the small sample sizes preclude a robust statistical analysis of seasonality on recapture rates, those months during which no recapture events occurred were drier months (Fig. 2). This might suggest that fly mobility, survival or loss of marking powder is influenced by rainfall or other climatic factors, but future studies with larger samples sizes across different seasons are needed to rigorously explore the importance of these factors.

We previously determined the fly species present inside the social groups of nonhuman primates in Kibale using

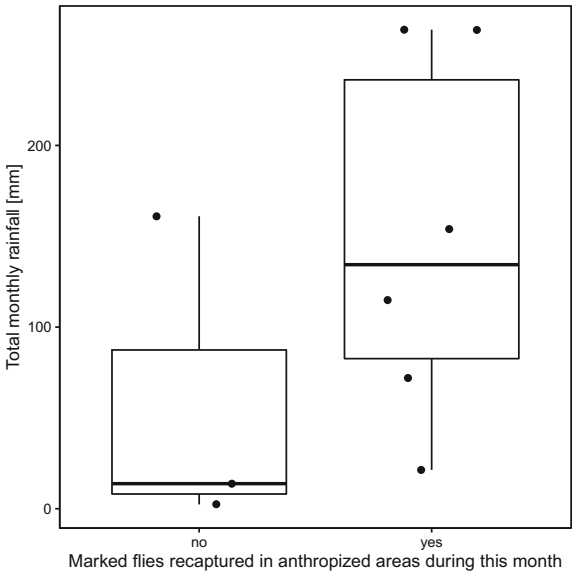

Figure 2. Box-and-whisker plot of monthly rainfall totals at the study site during months during which a recapture event occurred in the anthropized areas, or a recapture event did not occur. Lower and upper hinges correspond to the 25th and 75th percentiles, respectively, while the middle horizontal line represents the median. The upper and lower whiskers extend to the largest and lowest values no more than 1.5 times the interquartile range away from the hinge. Raw data are plotted in solid points.

soup metabarcoding of a fragment of the mitochondrial gene, cytochrome oxidase C subunit 1 (COI), detecting flies in the families Calliphoridae, Sarcophagidae and Muscidae (Gogarten et al. 2022). In these previous experiments, we removed a leg from flies captured in groups of six nonhuman primate species (N = 575 fly legs) and legs were pooled by nonhuman primate species and homogenized with a Tissuelyser II (Qiagen) and DNA extracted with the GeneMATRIX Stool DNA Purification Kit (Roboklon). To explore whether the same fly species were present outside nonhuman primate groups, we had also homogenized fly legs from the same flies captured 500 m outside groups (N_{files} = 575, N_{pools} = 6) and included a pool of 100 fly legs from flies captured in the Volkspark Rehberge, Berlin, Germany and an extraction blank as controls.

To explore the potential disease risk posed by the movement of flies from primate groups into anthropized, we molecularly characterized the eukaryotic parasites detected on flies captured in nonhuman primate groups. We performed soup metabarcoding to detect eukaryotic parasites, using the same pools of fly-leg extracts described above (for details of the extraction methods, see: Gogarten et al. 2022). Specifically, we applied a PCR system targeting the 18S rRNA of eukaryotic parasites (methods described in detail here: Maritz et al. 2017; Amaral-Zettler et al. 2018); we modified the protocol by using the two universal primers Euk 1391F and EukBr with nextera specific overhangs to amplify the V9 variable region of the 18S rRNA of eukaryotic parasites (Gohl et al. 2016). We then prepare amplicons for sequencing with a second PCR to append sequencing adapters and sample specific indexes. We included three negative controls and include one extraction blank with the PCR. Cycling conditions were 98°C for 5 min, 25 cycles of 98°C for 20 s, 65°C for 15 s, 57°C for 30 s and a final step of 72°C for 10 min (modified from Maritz et al. (2017). Products were visualized on 1.5% agarose gels and cleaned using AMPure XP Beads and pools uniquely dual indexed using the Nextera XT Index kit and sequenced on an Illumina NextSeq 500 with a mid-output kit v.2 and 2 × 150 cycles.

We removed primers using cutadapt (v 2.1: Martin 2011) and filtered reads using the DADA2 pipeline (Callahan et al. 2016) and assigned them taxonomically using the RDP naïve Bayesian classifier algorithm coupled with the PR2 training database (v 4.12.0: Wang et al. 2007). Poor read quality for the second read precluded its use in the analysis. To consider only parasites relevant to primate health, we considered reads assigned to families that in-

clude known primate parasites (following: Gogarten et al. 2020). We did not identify any such read in the negative controls (i.e., the extraction blank, the no template controls or the Berlin flies), but we detected reads belonging to three families of eukaryotic primate parasites in the flies captured in primate groups: Blastocystidae, Entamoebidae and Vahlkampfiidae (Fig. 3). In addition, from flies captured within the forest but outside primate groups, we detected parasites belonging to the same parasite families and the family Trypanosomatidae (no reads from flies of the Glossina genus were detected in the soup metabarcoding effort describing this fly community; Figure 3; Gogarten et al. 2022). While we here described the parasites carried by nonhuman primate-associated flies in Kibale, a clear limitation is that we did not collect and export the 19 primate-associated flies that we recaptured in anthropized areas, which precluded a determination of their species or the parasites these particular flies carried. This is an important area of future research and we encourage future research to explore which particular nonhuman primate-associated fly species tend to move into anthropized areas and the particular disease risk these specific fly species pose. Future research could provide an in-depth understanding of the individual fly species carrying particular pathogens and these species' particular behavior and interactions with primates that the current study could not provide. Furthermore, the detection of parasite DNA does not prove infectivity of these parasites and future studies are needed to explore the actual disease risk posed by the pathogens found in and on nonhuman primate-associated flies.

Vector-borne diseases appear to be emerging at an increasing rate, comprising a disproportionate share of emerging infectious diseases, particularly in Africa (Swei et al. 2020). The close evolutionary relationship and resultant similar physiology of nonhuman primates and humans make nonhuman primates a likely source for the zoonotic transmission of pathogens (Gillespie et al. 2008; Calvignac-Spencer et al. 2012, 2021). At the same time, human pathogens have shown their potential to cause mortality in nonhuman primate populations (Köndgen et al. 2008). Areas of between-species transmission are predicted to be highest around the forests of central and west Africa, where humans often come into contact with wild primates; contact between wildlife and humans is expected to rise as human populations continue to grow and habitat fragmentation increases (Pedersen and Davies 2009). Early studies of *Escherichia coli* bacteria in humans and primates in anthropically disturbed areas at the forest edge of Kibale National Park may suggest regular bidirectional, interspecific bacterial transmission (Goldberg et al. 2008), though more rigorous methods (e.g., phylogenomic analyses) are needed to confirm this hypothesis. With this study, we suggest that flies, which serve as mechanical vectors for infectious agents, require further consideration as vectors between human and wildlife populations. Understanding factors that facilitate the movement of flies across the human–wildlife interface may ultimately enable the implementation of mitigation strategies such as the construction and donation of latrines to people living on the edges of parks.

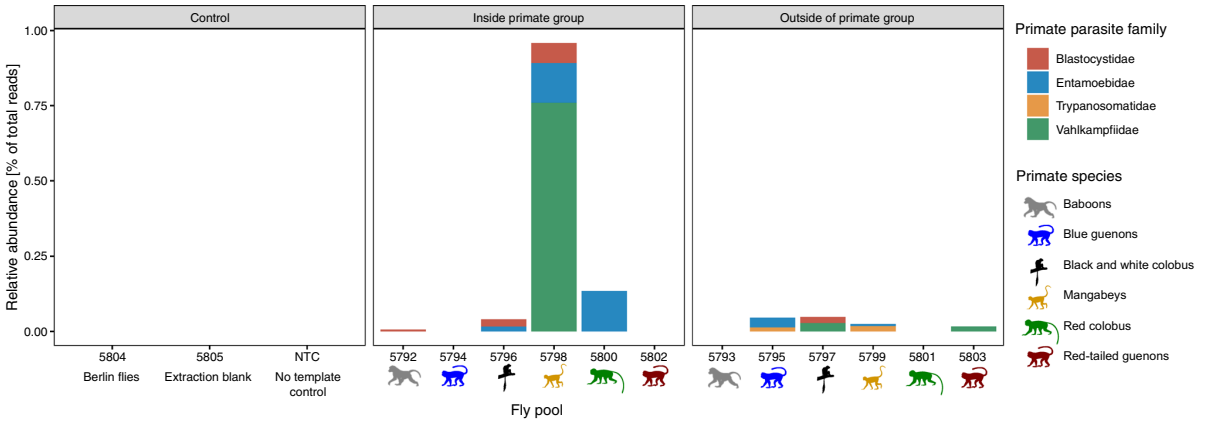

Figure 3. Relative abundance of primate parasite families detected in flies captured inside and outside primate groups and for the control experiments. Monkey silhouettes of different colors indicate the primate species from which flies were collected. The color of the solid bars indicates the primate parasite family detected. Numbers correspond the fly pool extract on which soup metabarcoding was performed (details available in: Gogarten et al. 2022) (Color figure online).

Acknowledgements

We are grateful to the motivated field teams of the Fish and Monkey Project for making this project possible, in particular, Michael Agaba, who performed the bulk of the field work associated with the project and Emmanuel Opito for his assistance with data entry. Ariane Düx provided valuable discussions and suggestions on the text. MJ was supported by the Bangabandhu Science and Technology Fellowship Trust. JFG was supported by funding from the Deutsche Forschungsgemeinschaft (DFG) Research Group "Sociality and Health in Primates" (FOR2136; CA 1108/3-1), as well as a DFG grant (GO 3443/1-1: Project number: 453352748). CAC and the Fish and Monkey Project were supported through National Geographic Society and the Leakey Foundation. CAC was supported by the Wilson Center. This project benefited immensely from synergies and discussions with the DFG project Great Ape Health in Tropical Africa: Emerging diseases, conservation and evolutionary perspectives on human health (GATECEP: LE1813/14-1).

Author Contributions

All authors designed the project. MJ and JFG performed the analysis. PAO and CAC organized and supervised the data collection in Kibale. JFG wrote the first draft of the manuscript and all authors contributed to writing and approved the final version and agreed to be accountable for all aspects of the work.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Data availability

Metabarcoding sequencing data are available as a Zenodo dataset: https://doi.org/10.5281/zenodo.7509929.

Declarations

Conflict of interest The authors declare no conflicting or competing interests.

Consent to Publish All authors gave final approval for publication.

Ethical approval Permission to conduct research on the flies associated with primates in Kibale National Park was given by the Uganda National Council for Science and Technology and the Uganda Wildlife Authority.

References

Adenusi AA, Adewoga TO (2013) Studies on the potential and public health importance of non-biting synanthropic flies in the mechanical transmission of human enterohelminths. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 107:812–818

Amaral-Zettler LA, Bauer M, Berg-Lyons D, Betley J, Caporaso JG, Ducklow W, Walters W (2018) EMP 18S Illumina amplicon protocol. *Protocols* io:1–8

Barnard C (1952) Yaws and Flies. Past and Present Opinions on the Role of Flies in the Transmission of Framboesia Tropica. *Journal of Tropical Medicine and Hygiene* 55:135–141

Calibeo-Hayes D, Denning SS, Stringham S, Guy JS, Smith LG, Watson DW (2003) Mechanical transmission of turkey coronavirus by domestic houseflies (*Musca domestica* Linnaeus). *Avian Diseases* 47:149–153

Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP (2016) DADA2: high-resolution sample inference from Illumina amplicon data. *Nature Methods* 13:581–583

Calvignac-Spencer S, Düx A, Gogarten JF, Leendertz FH, Patrono LV (2021) A great ape perspective on the origins and evolution of human viruses. In: *Advances in Virus Research*, Academic Press, pp 1–26

Calvignac-Spencer S, Leendertz SAJ, Gillespie TR, Leendertz FH (2012) Wild great apes as sentinels and sources of infectious disease. *Clinical Microbiology and Infection* 18:521–527

Chapman CA, Galán-Acedo C, Gogarten JF, Hou R, Lawes MJ, Omeja PA, Sarkar D, Sugiyama A, Kalbitzer U (2021) A 40-year evaluation of drivers of African rainforest change. *Forest Ecosystems* 8:66

Clavel A, Doiz O, Morales S, Varea M, Seral C, Castillo FJ, Fleta J, Rubio C, Gómez-Lus R (2002) House fly (*Musca domestica*) as a transport vector of *Cryptosporidium parvum*. *Folia Parasitologica* 49:163–164

Forsey T, Darougar S (1981) Transmission of chlamydiae by the housefly. *British Journal of Ophthalmology* 65:147–150

Gillespie TR, Nunn CL, Leendertz FH (2008) Integrative approaches to the study of primate infectious disease: Implications for biodiversity conservation and global health. *American Journal of Physical Anthropology* 137:53–69

Gogarten JF, Calvignac-Spencer S, Nunn CL, Ulrich M, Saiepour N, Nielsen HV, Deschner T, Fichtel C, Kappeler PM, Knauf S, Müller-Klein N, Ostner J, Robbins MM, Sangmaneedet S, Schülke O, Surbeck M, Wittig RM, Sliwa A, Strube C, Leendertz FH, Roos C, Noll A (2020) Metabarcoding of eukaryotic parasite communities describes diverse parasite assemblages spanning the primate phylogeny. *Molecular Ecology Resources* 20:204–215

Gogarten JF, Düx A, Mubemba B, Pléh K, Hoffmann C, Mielke A, Müller-Tiburtius J, Sachse A, Wittig RM, Calvignac-Spencer S, Leendertz FH (2019) Tropical rainforest flies carrying pathogens form stable associations with social nonhuman primates. *Molecular Ecology* 28:4242–4258

Gogarten JF, Jahan M, Calvignac-Spencer S, Chapman CA, Goldberg TL, Leendertz FH, Rothman JM (2022) The cost of living in larger primate groups includes higher fly densities. *Ecohealth*

Gohl DM, Vangay P, Garbe J, MacLean A, Hauge A, Becker A, Gould TJ, Clayton JB, Johnson TJ, Hunter R, Knights D, Beckman KB (2016) Systematic improvement of amplicon marker gene methods for increased accuracy in microbiome studies. *Nature Biotechnology* 34:942–949

Goldberg TL, Gillespie TR, Rwego IB, Estoff EL, Chapman CA (2008) Forest fragmentation as cause of bacterial transmission among nonhuman primates, humans, and livestock, Uganda. *Emerging Infectious Diseases* 14:1375–1382

Graczyk TK, Knight R, Gilman RH, Cranfield MR (2001) The role of non-biting flies in the epidemiology of human infectious diseases. *Microbes and Infection* 3:231–235

Greenberg B (1973) *Flies and disease. Vol. II. Biology and disease transmission*

Hoffmann C, Zimmermann F, Biek R, Kuehl H, Nowak K, Mundry R, Agbor A, Angedakin S, Arandjelovic M, Blankenburg A, Brazolla G, Corogenes K, Couacy-Hymann E, Deschner T, Dieguez P, Dierks K, Düx A, Dupke S, Eshuis H, Formenty P, GinathYuh Y, Gogarten JF, Goedmakers A, Granjon A, McGraw S, Grunow R, Hart J, Jones S, Junker J, Kiang J, Langergraber K, Lapuente J, Lee K, Leendertz SAJ, Leinert V, Löhrich T, Marrocoli S, Mätz-Rensing K, Meier A, Merkel K, Metzger S, Murai M, De Nys HM, Sachse A, Schenk S, van Schijndel J, Thiesen U, Ton E, Wieler LH, Boesch C, Klee SR, Wittig RM, Calvignac-Spencer S, Leendertz FH (2017) Persistent anthrax as a major driver of wildlife mortality in a tropical rainforest. *Nature* 548:82–86

Khamesipour F, Lankarani KB, Honarvar B, Kwenti TE (2018) A systematic review of human pathogens carried by the housefly (*Musca domestica* L.). *BMC Public Health* 18:1–15

Knauf S, Raphael J, Mitjà O, Lejora IA, Chuma IS, Batamuzi EK, Keyyu JD, Fyumagwa R, Lüert S, Godornes C (2016) Isolation of *Treponema* DNA from necrophagous flies in a natural ecosystem. *EBioMedicine* 11:85–90

Köndgen S, Kühl H, N'Goran PK, Walsh PD, Schenk S, Ernst N, Biek R, Formenty P, Mätz-Rensing K, Schweiger B (2008) Pandemic human viruses cause decline of endangered great apes. *Current Biology* 18:260–264

Lamboen W (1936) The experimental transmission to man of *Treponema pertenue* by the fly *Musca sorbens*, WD. *Journal of Tropical Medicine and Hygiene* 15:235–239

Maritz JM, Rogers KH, Rock TM, Liu N, Joseph S, Land KM, Carlton JM (2017) An 18S rRNA Workflow for Characterizing Protists in Sewage, with a Focus on Zoonotic Trichomonads. *Microb Ecol* 74:923–936

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *Embnet. Journal* 17:10–12

Pedersen AB, Davies TJ (2009) Cross-Species Pathogen Transmission and Disease Emergence in Primates. *Ecohealth* 6:496–508

Refisch J, Koné I (2005) Market hunting in the Tai Region, Cote d'Ivoire and implications for monkey populations. *International Journal of Primatology* 26:621–629

Rosef O, Kapperud G (1983) House flies (*Musca domestica*) as possible vectors of *Campylobacter fetus* subsp. *jejuni*. *Applied and Environmental Microbiology* 45:381–383

Stoffolano JG (2022) Synanthropic flies—a review including how they obtain nutrients, along with pathogens, store them in the crop and mechanisms of transmission. *Insects* 13

Swei A, Couper LI, Coffey LL, Kapan D, Bennett S (2020) Patterns, Drivers, and Challenges of Vector-Borne Disease Emergence. *Vector-Borne and Zoonotic Diseases* 20:159–170

Wang Q, Garrity GM, Tiedje JM, Cole JR (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology* 73:5261–5267

Yu DW, Ji Y, Emerson BC, Wang X, Ye C, Yang C, Ding Z (2012) Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods in Ecology and Evolution* 3:613–623

Linking statement 2

In my first chapter, I successfully employed fly iDNA to contribute to our understanding of the disease ecology and distribution pattern of the pathogen Bcbva, along the different habitats spanning from rainforest to the surrounding villages. In addition, I aimed to use metabarcoding of fly iDNA to reconstruct the communities of flies and mammals for investigating potential links between their presence and the detection of Bcbva. The study was published in *Environmental DNA*. In my second chapter, published in *EcoHealth*, I presented another fly iDNA based biomonitoring study where instead of extracting single fly or small pools of flies, as has been the predominant approach in the field, I highlighted the effectiveness of using bulk fly leg pools to investigate the diversity of eukaryotic parasites carried by primate-associated flies in the environments surrounding Kibale National Park, Uganda.

Considering the overall promise of fly iDNA for host and pathogen studies in tropical and terrestrial ecosystems, the successful implementation of mass fly leg pooling in Chapter two highlights an important avenue to optimize the scalability and usability of fly iDNA based monitoring by potentially reducing extraction costs and resources needed. With this aim, I explored the feasibility of two different bulk fly extraction schemes for monitoring mammal diversity. I do this by working with flies collected on a large geographic scale in five different countries of sub-Saharan Africa. The first method involved destructive extraction of large pools of fly legs, while the second method involved a non-destructive extraction method of large numbers of entire fly bodies. The findings of this chapter suggest that for mammal detection, mass DNA extraction of large numbers of flies in a single extraction pool using a non-digestion buffer, recovers similar numbers of mammal detections compared to the previous study by Gogarten et al. (2020), paving the way for the adoption of fly iDNA for terrestrial biomonitoring at large scales for monitoring both wildlife and pathogen diversity.

3.3 Chapter 3

An exploration of methods to allow for the bulk extraction of fly iDNA allowing for scalable biodiversity monitoring

Mueena Jahan[1,2,3], Christina Lynggaard[4,5], Fabian H. Leendertz[1,2], Jan F. Gogarten[2,4,6,7*], Sébastien Calvignac-Spencer[2,4,6*]

[1] Department of Ecology and Emergence of Zoonotic Diseases, Helmholtz Institute for One Health, Helmholtz-Centre for Infection Research (HZI), Greifswald, Germany

[2] Epidemiology of Highly Pathogenic Organisms, Robert Koch Institute, Berlin, Germany

[3] Department of Microbiology and Public Health, Bangabandhu Sheikh Mujibur Rahman Agricultural University, Gazipur, Bangladesh

[4] Department of Pathogen Evolution, Helmholtz Institute for One Health, Helmholtz-Centre for Infection Research (HZI), Greifswald, Fleischmannstraße 42, 17489 Greifswald, Germany

[5] Section for Evolutionary Genomics, The GLOBE Institute, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen K, Denmark

[6] Viral Evolution, Robert Koch Institute, Berlin, Germany

[7] Applied Zoology and Nature Conservation, University of Greifswald, Greifswald, Germany

[*] Jan F. Gogarten and Sébastien Calvignac-Spencer should be considered joint senior author.

[#]Correspondence: jan.gogarten@helmholtz-hioh.de and sebastien.calvignac-spencer@helmholtz-hioh.de

Keywords

Bulk pooling, Environmental DNA (eDNA), invertebrate-derived DNA (iDNA), mammals, non-destructive DNA extraction, terrestrial biomonitoring

Author contributions

MJ, SCS, and JFG contributed to the conception or design of the study and MJ, CL, and JFG prepared the first draft of the manuscript. All authors contributed to the acquisition, analysis, or interpretation of the data, as well as contributed to editing and shaping the final manuscript.

Acknowledgments

We are grateful to the Pan African Programme: The Cultured Chimpanzee (PanAf), for their collection of flies during their extensive field efforts. In particular, for their support in organizing the missions and conceiving and funding the program, as well carrying out the field work, we thank the following team members who will be involved in the final publication of this this chapter: Mimi Arandjelovic, Hjalmar Kühl, Constanze Hoffmann, Fee Zimmermann, Andreas Sachse, Kevin Merkel, Paula Dieguez, Anthony Agbor, Samuel Angedakin, Gregory Brazzola, Sorrel Jones, Kevin E. Langergraber, Kevin Lee, Sergio Marrocoli, Mizuki Murai, and Volker Sommer.

For their support and permission to conduct this study, we would like to thank the following collaborating ministries and organizations: Ministere de l'Agriculture de l'Elevage et des Eaux et Forets, Guinea; Forestry Development Authority, Liberia; National Park Service, Nigeria; Direction des Eaux, Forêts et Chasses, Senegal; Makerere University Biological Field Station, Uganda; National Council for Science and Technology, Uganda; and Wildlife Authority and the National Forestry Authority, Uganda. We also thank the following nongovernmental organizations for their support: Wild Chimpanzee Foundation, Guinea; Wild Chimpanzee Foundation, Liberia; Budongo Conservation Field Station, Uganda; and Ngogo Chimpanzee Project, Uganda. We are extremely grateful for Christophe Boesch's vision and dedication in establishing the PanAf program. We also thank Katherine Corogenes, Dervla Dowd, Henk Eshuis, Vincent Lapeyre, Yasmin Moebius, Geoffrey Muhanguzi, Erin G. Wessling, and Klaus Zuberbuehler for their help with data collection and field logistics.

We also thank the Sequencing Unit of the Robert Koch Institute for their assistance with this project. MJ was funded by the Bangabandhu Science and Technology Fellowship Trust, Ministry of Science and Technology, Government of the People’s Republic of Bangladesh. JFG was supported by the DFG grant “The ecology and evolution of primate phageomes” (GO 3443/1-1: Project number: 453352748).

Data availability statement

All metabarcoding sequences generated for this study have been uploaded to Zenodo (https://zenodo.org/record/xxx), with the logic that assignment to a particular organism is always uncertain and should not supersede existing sequences linked to a particular host in available sequence databases.

ABSTRACT

Background

Metabarcoding of invertebrate-derived DNA (iDNA) has proven an excellent tool for assessing terrestrial mammal diversity. Fly iDNA-based monitoring can detect a wide range of mammals, including rare, small, and cryptic species. However, a major constraint to the widespread adoption of fly iDNA-based mammal monitoring has been the significant costs associated with sample processing.

Aims and Methods

Here, we explored the effectiveness of both destructive and non-destructive bulk DNA extraction approaches for gaining access to fly iDNA for mammal biomonitoring. One extraction method involved removing a single leg from many flies, pooling the legs, and destructively extracting DNA from the pool. The other method involved non-destructively extracting DNA from a large pool of entire flies. We tested these methods on flies collected at eight sites, representing three different habitats across five countries in sub-Saharan Africa ($N_{flies/site}$ = 105; $N_{flies\ total}$ = 840). We compared the mammal species detected with these two approaches with a previous study using fly iDNA in these ecosystems, extracted destructively using small pools of entire flies ($N_{flies\ per\ pool}$ = 7; $N_{flies\ per\ site}$ = 105 flies).

Results

The non-destructive extraction method using entire files detected a greater number of mammals ($N_{species\ detections\ total}$ =59; $\bar{x}_{species/site}$ =7.4; range = 1 to 12) than the destructive extraction of pools of fly legs ($N_{species\ detections\ total}$ = 15; $\bar{x}_{species/site}$ =1.9; range = 1 to 3). Mass DNA extraction of large numbers of flies in a single extraction using a non-digestion buffer resulted in a similar number of mammal detections compared to destructive DNA extraction from small pools of entire flies ($N_{species\ detections\ total}$ = 67; $\bar{x}_{species/site}$ = 8.4; range = 5 to 15).

Conclusion

Our findings indicate that mass pooling entire fly bodies and subsequently using a non-digestive buffer for DNA extraction has the potential to streamline fly iDNA extraction efforts. This, in turn, paves the way for the widespread use of fly iDNA in large-scale terrestrial biomonitoring efforts.

INTRODUCTION

Massive declines in the abundance and diversity of animals on a global scale are a result of anthropogenic disturbance, with particularly pronounced impacts in tropical terrestrial ecosystems. Rapid terrestrial biodiversity loss could alter ecosystem dynamics, disrupting the intricate web of interactions within ecosystems. For example, when certain keystone species disappear this may lead to cascading impacts on the entire ecosystem, while a loss of diversity might change infectious disease dynamics and change human disease risk (Hagen et al. 2012; Keesing et al. 2010). Wildlife monitoring can offer insights into real-time changes within ecosystems and thereby shape decision-making regarding conservation efforts. However, monitoring wild animal diversity in resource constrained tropical biodiversity hotspots still poses significant challenges, despite these areas being the ones undergoing the most rapid environmental change. Therefore, tools allowing for systematic monitoring of wildlife species are critically needed (Butchart et al. 2010), particularly tools that simplify and streamline the assessment of biodiversity, reduce costs as well as the required level of expertise (Bohmann et al. 2013; Calvignac-Spencer et al. 2013).

In this context, environmental DNA (eDNA; genomic DNA from different organisms shed into the environment) is a promising approach that facilitates biodiversity assessments (Taberlet et al. 2018; Bohmann et al. 2014). Metabarcoding of eDNA has proved to be particularly effective for the assessment of vertebrate and invertebrate species diversity in marine ecosystems (Valentini et al. 2016; Rees et al. 2014; Goldberg et al. 2011). However, for terrestrial fauna, eDNA-based monitoring is notably more challenging due to the lack of suitable sampling substrates or the difficulty in obtaining them. Despite these challenges, various substrates have been explored as sources of terrestrial eDNA. Indeed, terrestrial mammal and vertebrate eDNA has been detected in various substrates, including soil (Leempoel et al. 2020; Andersen et al. 2012), forest water (Abrams et al. 2019; Ushio et al. 2017), in airborne particles (Clare et al. 2022; Lynggaard et al. 2022a; Lynggaard et al. 2022b), swabs from herbaceous vegetation for small mammal detection (Lynggaard et al. 2023; Lyman et al. 2022), flowers (Newton et al. 2023), and rainwash (Macher et al. 2023). To date, none of these substrates has yet emerged as a go to solution for wildlife managers and conservationists, akin to what water has become for biomonitoring efforts in aquatic systems.

An intriguing approach is to explore substrates that are more closely linked with the wildlife of interest. For example, the collection of arthropod eDNA at flowers was used to reveal those arthropods that interact with those specific plants (Johnson et al. 2023). For terrestrial vertebrates, a promising source of high quality eDNA is invertebrate derived DNA (iDNA) collected from organisms that consume or interact with vertebrate DNA as part of their lifecycle (Calvignac-Spencer et al. 2013). Several studies have explored the potential of both

hematophagous and non-hematophagous invertebrates for detecting vertebrate terrestrial biodiversity. Indeed, hematophagous insects such as mosquitoes (Grubaugh et al. 2015), sand flies (Kocher et al. 2017), biting midges (Cutajar & Rowley, 2020; Lassen et al. 2011), leeches (Abrams et al. 2019; Schnell et al. 2018; Schnell et al. 2012), and ticks (Gariepy et al. 2012), as well as non-hematophagous scavengers like dung beetles (Gollan et al. 2013); carrion flies (flies that feed on animal carcasses) (Owings et al. 2019; Rodgers et al. 2017; Calvignac-Spencer et al. 2013),or even bulk arthropod samples (Lynggaard et al. 2019) have been employed effectively for unraveling terrestrial biodiversity along environmental gradients (Massey et al. 2022), (Gogarten et al. 2020) and between local habitats (Jahan et al. 2023; Srivathsan et al. 2022). In an impressive effort that scaled up such efforts, 30,468 leaches were collected and used to assess the efficacy of protected areas by gauging the distribution of vertebrate communities using leech iDNA, which allowed for multi-species occupancy modelling (Ji et al. 2022). While certain sources of iDNA (e.g. leeches) can be trapped opportunistically (Schnell et al. 2012) and others often exhibit certain level of host preferences (e.g. mosquitoes) (Lyimo and Ferguson 2009), carrion flies are very easy to trap, and show no clear specific host preferences (Calvignac-Spencer et al. 2013).

Metabarcoding of carrion fly derived iDNA has been used as an effective tool for detecting DNA from a wide array of mammal and vertebrate species (Rodgers et al. 2017; Lee et al. 2016; Calvignac-Spencer et al. 2013). These include animals of varying sizes, as well as those that inhabit remote areas with low population densities in tropical and temperate environments. Direct comparisons of fly iDNA with camera trapping suggest that fly iDNA detects smaller species that often go unnoticed by camera trapping or direct observation at both small (Rodgers et al. 2017; Lee et al. 2016) and large spatial scales (Gogarten et al. 2020). Intruigingly, targeted species detection approaches using iDNA appear more sensitive than those targeting broader taxonomic groups, suggesting fly iDNA has an untapped potential for detecting particular terrestrial species of interest (e.g. invasive species, endangered species) (Schubert et al. 2015). Fly iDNA has also demonstrated promise in monitoring mammal species within urban settings in a temperate climate (Hoffmann et al. 2018), showcasing the versatility of fly iDNA surveillance in a diversity of ecosystems. Fly iDNA has even been used to reconstruct mitogenomes of mammals, presenting an intriguing opportunity to evaluate genetic diversity at the population level using iDNA (Danabalan et al. 2023).

One of the most important aspects of fly iDNA based biomonitoring is that extremely large numbers of flies can be easily collected by non-experts (e.g., a single fly-trap set for 20 minutes in Taï National Park collected over 700 flies (Gogarten et al. 2019)). However, all of the aforementioned studies using fly iDNA have worked with individual flies or relatively small batches of flies (typically <10 flies / pool), which makes scaling up to 10’s or even 100’s of thousands of flies for large scale biomonitoring challenging. With the aim to increase

throughput, Srivathsan et al. (2022) captured live flies and used an extraction free fly iDNA metabarcoding protocol to explore a vertebrate community in large swamp forest remnant in Singapore, eliminating extraction reagent costs and reducing the time needed to carry out experiments. Indeed, Srivathsan et al. (2022) showed that by trapping a single fly alive in a tube, that they could then easily collect the fly feaces and regurgitate and directly proceed with a PCR after resolubilizing the fly biproducts; this allowed them to identify vertebrate taxa from 60% of the collected flies. While striking in its simplicity the approach required handling of individual live flies and putting them into separate tubes, limiting the number of flies that could be quickly processed in the field. Thus, despite the potential for the collection of extremely large numbers of flies, the challenges associated with the existing methods used for DNA extraction and sample processing has precluded the adoption of fly iDNA for terrestrial biodiversity monitoring at large scales.

There is thus a clear need for method optimization to allow this approach to be more widely adopted and used at scale. Indeed, different efforts have been made to streamline methods for bulk extraction of invertebrates for terrestrial biomonitoring, such as the use of a non-destructive extraction protocol developed by Gilbert et al. (2007) to extract DNA from bulk arthropod samples to explore vertebrate diversity in a tropical forest (Lynggaard et al. 2019), as well as non-destructive extraction of large pools of leeches (Ji et al. 2022). In addition, DNA extraction of large pools of just individual fly legs has been used for characterizing eukaryotic primate parasites in and around rainforest ecosystems (Jahan et al. 2022). Nevertheless, this last approach has not yet been explored for terrestrial mammal species detection, which we attempted to do here. We explore the potential of two bulk DNA extraction approaches to enhance the sampling throughput and expedite the analysis of fly iDNA and compare them to the extraction method used by Gogarten et al. (2020). Specifically, we compare the following extraction methods: 1) destructive DNA extraction from large pools of fly legs, 2) non-destructive DNA extraction from pools of entire fly bodies and 3) ‘classic’ destructive DNA extraction of smaller pools of entire fly bodies used by Gogarten et al. (2020) and numerous other studies (Danabalan et al. 2023; Hoffmann et al. 2018; Rodgers et al. 2017; Calvignac-Spencer et al. 2013). The method used by Gogarten et al. (2020) involves DNA extraction from a large number of smaller pools, which leads to a much higher extraction material and time costs compared to the two bulk extractions methods explored in this study. Therefore, our aim was to explore if these bulk methods that require only a single extraction pool are as efficient for terrestrial mammal detection as the method using more pools.

MATERIALS AND METHODS

Study sites and sample collection

To allow a comparison to the results obtained in Gogarten et al. (2020), a subset of the carrion flies collected in the same campaign in 2013 were chosen for this study (see Gogarten et al. (2020) for details about this fly collection effort; Figure 1). In short, carrion flies were collected in eight different sites located in five countries depicting three habitats in tropical sub-Saharan Africa. Among these, two represent forest-savannah ecosystems (Gashaka Gumti National park in Nigeria and Fouta in Guinea), one a savannah ecosystem (Kayan in Senegal) and the rest rainforest ecosystems: three in east African rainforests (Budongo forest, Bwindi Impenetrable National Forest and Ngogo East in Kibale National Park, all in Uganda) and two in West African rainforests (Grebo National Forest and East Nimba National reserve in Liberia) (Gogarten et al. 2020). For East Nimba National reserve, all collected flies had been destructively analyzed by Gogarten et al. (2020) and none were left for this study. Here, we selected flies which were trapped in parallel in 2013 in Taï National park, Côte d'Ivoire, which also represents a West African rainforest ecosystem. We thus compare the findings from the previous study in Nimba, with the bulk extraction methods applied to Täi dataset; at all other sites, the comparison of methods was done on flies collected in the same locations. Gogarten et al. (2020) analyzed 105 flies at each site (with the exception of East Nimba Nature Reserve for which only 49 were collected and analyzed), which were then divided into 15 pools of 7 flies each (with the exception of East Nimba Nature Reserve for which only 7 fly pools were analyzed). Thus, in total, 784 flies were processed for further molecular analyses. Here, we selected 105 flies collected at each site, which were processed in a single pool, meaning a total of 840 flies were processed for this study.

DNA extraction

DNA was extracted using two extraction methods. For the first extraction method, at each site a single leg was separated from each of the 105 flies and pooled into one tube prior to extraction (Figure 1). This resulted in a total of eight leg pools, one for each site. Thereafter, DNA was extracted from these leg pools using the GeneMatrix Stool DNA purification kit (Roboklon, Berlin, Germany) as described in (Jahan et al. 2023). In brief, fly legs were moved into a bead tube and sliced into smaller pieces with autoclaved scissors and then homogenized using the Tissue lyser II (Qiagen). After centrifugation, supernatants were subsequently treated according to the manufacturer's protocol. Concentration of fly legs DNA were measured with Qubit 3 fluorometer and high sensitivity dsDNA assay kit (Invitrogen by Thermo Fisher Scientific).

For the second extraction method, the 105 fly bodies (missing the leg that was used for the first method) were pooled together into a 50ml falcon tube and adequate amount of digest

buffer was added so that fly masses were covered (see Nielsen et al. (2019) for details about the digestion buffer). Samples were then placed on a shaking incubator at 56℃ for overnight incubation and then samples were briefly centrifuged to allow the fly masses to sediment at the bottom of the tube. After that, 200 µl of DNA digest from each of the bulk fly pool sample were transferred into the Eppendorf tube (2ml) and further purified using the QiaQuick PCR purification kit (Qiagen), following the manufacturer's protocol with a modification at the final step as detailed in Lynggaard et al. (2019). After collecting the DNA digest, each of the bulk fly pool sample were removed from the buffer and placed in 100% ethanol for 2-4 hours to stop digestion and then samples were returned to the collection tube, replenished with 70% ethanol, and preserved for potential future analysis.

Laboratory work followed laboratory safety protocols including surface cleaning, disinfection, and exposure to UV light for 15 minutes. As an additional control in the experiment, 105 flies collected from different parks in Berlin, Germany, were also processed. As a positive control for the fly leg-based and for the fly body-based extraction methods, legs and fly body parts from these flies were processed as the ones from Africa. Extraction blanks were also processed during both extraction methods as a negative control (n=2 per method).

Mammal metabarcoding

To reconstruct the mammal species diversity at each site, metabarcoding was then performed using the fly DNA extracts and a three steps PCR amplification to minimize the amplification bias, as suggested by Hoffmann et al. (2017). For the first set of mammal PCRs, around 130 bp of the 16S mitochondrial DNA marker was PCR amplified using two universal primers,16Smam1(5′-CGGTTGGGGTGACCTCGGA-3′) and 16Smam2 (5′-GCTGTTATCCCTAGGGTAACT-3′) (Calvignac-Spencer et al. 2013) along with two sets of blocking oligonucleotide primers for human and pig to minimize the potential amplification of contamination, 16S mam_blk hum 3 (5′- CGGTTGGGGCGACCTCGGAGCAGAACCC- 3′) and 16Smam_blkpig (5′- CGGTTGGGGT GACCT CGG AGT ACA AAAAAC- 3′) as described in Calvignac-Spencer et al. (2013). During PCR amplification of the positive control, two additional sets of blocking primers were used to reduce the amplification of potential contamination from dog and cattle,16Smam_blkcan1(5′-AGGTTGGGGTGACCTCGGAATATAAAAAAACTC-3′) and 16Smam_blkbos2(5′_CGGTTGGGGTGACCTCGGAGAATAAAAAATCCTCCGAGCGATTTTAAAG- 3′) respectively. All PCRs were performed with four replicates for all the fly extracts including the negative and positive controls.

Each PCR was carried out in a total volume of 25 µl that consisted of 4 µl DNA template, 2.5 mM dNTP (replaced by dUTP), 50 mM MgCl2, 10 µM of each primer, 10 µM of two blocking

primers (human and pig, and dog and cattle for the controls), 0.3 U Amperase® uracil N-glycosylase (Invitrogen), 1.25 U Platinum® Taq Polymerase (Invitrogen), 2.5 µl of 10× PCR Buffer (Invitrogen), and nuclease free molecular grade water. Cycling conditions were as follows: 45°C for 7 min, 95°C for 10 min, 42 cycles of 95°C for 30 s, 64°C for 30 s, 72°C for 1 min, followed by elongation at 72°C for 10 min. PCR amplified products were then visualized on 1.5% agarose gel for the expected size bands. All the negative controls were negative on the gel, but were considered for the further analyses as control. All the amplicons with the positive bands were further gel excised under the UV trans-illuminator and purified using the purelink™ quick gel extraction kit (Invitrogen, Thermo fisher Scientific). Gel purified PCR amplicons were then processed for next step PCR. For the second set of PCR, the total volume for each reaction was 25 µl and consisted of 5 µL of gel purified 16S PCR amplicons, 2.5 mM dNTP, 4 mM MgCl2, 10 µM of each fusion primer (16Smam primer appended with overhang Illumina specific adapter sequence), 1.25 U Platinum® Taq Polymerase (Invitrogen), and 2.5 10× PCR Buffer (Invitrogen). Cycling conditions were as follows: 95°C for 5 min, 15 cycles of 95°C for 30 s, 64°C for 30 s, 72°C for 1 min, and a final elongation of 72°C for 10 min. These amplified products were cleaned using paramagnetic beads with 1.8 µL × MagSi NGSPREP Plus beads to 1 µL amplicon ratio and eluted in 17.5 µL TET buffer and then dual indexed using Nextera Index kit followed by the index PCR amplification. The protocol in brief, each of the 12.5 µL PCR reaction mixture included 1.25 µL of bead purified PCR amplicons, 6.25 µL 2× KAPA HiFi HotStart ReadyMix (peqLab), 1.25 µL of each Nextera XT index primer (Illumina) and 2.5 µL of nuclease free water. Cycling conditions were 95°C for 3 min, 8 cycles of 95°C for 30 s, 55°C for 30 s, 72°C for 30 s, and a final elongation step of 72°C for 5 min. The amplicons from index PCR were again purified using 1.8 µL × MagSi NGSPREP Plus beads to 1 µL amplicon ratio and eluted in a 17.5 µL TET buffer. The indexed amplicon libraries were then quantified using a Qubit 3 and the high sensitivity dsDNA assay kit (Invitrogen) and sequenced on an Illumina Nextseq 500 with a mid-output kit v.2 and 2 × 150 cycles.

In silico analyses of mammal metabarcoding

The in silico analyses of sequences generated, was performed following the pipeline outlined by Gogarten et al. (2020). In brief, paired end raw reads were joined using the *illuminapairedend* command of the OBITools package (v1.2.13), setting the minimum alignment score to 40 and excluding reads that do not overlap (Boyer et al. 2016). The *ngsfilter* command in OBITools was then employed to remove primer sequences, followed by quality trimming using Trimmomatic (v0.36), using a minimum quality score of 30 over a sliding window of four bases, as well as a leading and trailing minimum quality score of 30 and a minimum surviving read length of 80 bp. Subsequently, remaining reads were subjected to de-replication using the OBITools *obiuniq* command. A reference database was constructed using the OBITools *ecoPCR* (v0.2) command for conducting an in silico PCR on all mammal

sequences present in GenBank (accessed Nov. 2021), allowing three mismatches between primers and reference sequences and a synthetic product length between 50 and 800 bp. The database was then employed to assign taxonomy to the remaining reads from our experiment using the OBITools *ecotag* command, with a minimum identity level set at 0.97. Downstream analyses were conducted based on sequences that were assignable to a mammal species. Three species were detected in either positive or negative controls (*Cercopithecus diana* in positive control; *Genetta thierryi* and *Cricetomys emini* in negative control), thus we excluded any reads that were assigned to these taxa from downstream analysis.

In addition, we observed some low-level detections of African mammals in the extraction blanks, so to consider a detection genuine, we applied a threshold of two times the maximum number of unexpected reads assigned to unexpected taxa detected in any of the controls, extraction blanks, or Berlin flies (i.e., 29 reads assignable to the species *Cercopithecus albogularis* in an extraction blank). Thus, we only considered a detection event of a species in a sample genuine if it had more than 58 reads. We also excluded detection events of domestic animal species that represent frequent laboratory contaminants, as well as any human detection events. In addition, to keep results comparable with Gogarten et al. (2020) we filtered the datasets by excluding detections of species not present in the country according to the IUCN Red List of Threatened Species (https://www.iucnredlist.org/) and the Global Biodiversity Facility (GBIF) list (https://www.gbif.org/); this helped remove nine detections that could have represented a closely related sister species that were not in the reference database or a sequencing/PCR artifact from an amplicon from a closely related species also detected in the library (accessed on 1st July 2023).

Statistical analysis and visualization

We compared the present dataset obtained from the bulk approaches, namely the bulk fly leg pool based destructive extraction method and the bulk fly body pool non-destructive method, with the previous mammal detection datasets derived from the entire fly pool destructive extraction method, that used multiple smaller fly pools (Gogarten et al. 2020). We performed an analysis of variance (ANOVA) to identify differences in species counts based on the three extraction methods. We the performed a Tukey's Honestly Significant Difference (HSD) post-hoc test using the *multcomp* package (Hothorn et al. 2008) to test which pair of extraction methods differed in their species counts.

To visualize the number of mammal species detected at each site with three different extraction methods, we constructed a box-whisker plot using the *ggplot2* R package (Kahle and Wickham 2013). All statistical analyses and visualization was performed in R v4.2.0 (R core team 2021)

RESULTS

For this study, we analyzed a total of 16 fly pools with 105 flies each. Eight of these pools were generated using the bulk fly leg pool based destructive extraction method. A further eight pools were generated using the remaining fly (without one leg) which were processed using the bulk fly body pool non-destructive method. In addition, one set of positive and negative controls was included in for each of the extraction methods. From the eight pools processed with the bulk fly leg pool based destructive method, it was possible to generate amplicons from all four PCR replicates from six pool extracts, but the fly leg pool replicates from Taï and Ngogo sites exhibited lower amplification success. It was possible to amplify mammal DNA from two PCR replicates from the Taï pool, whereas this was only possible for one PCR replicate from the Ngogo pool. In contrast, each replicate from the eight bulk fly body pool extracts showed a successful amplification of mammal DNA using the bulk fly body pool non-destructive method.

Mammal community detected

As also found in Gogarten et al. (2020), the majority of mammals were detected in Grebo, Liberia. However, no mammal taxa were detected from Ngogo, Uganda and Täi national park, Côte d'Ivoire using the bulk fly leg pool based destructive method (Table 1). When combining the results from the two extraction methods tested, we detected a total of 74 species in all eight sites that encompassed 20 families belonging to eight orders (Artiodactyla, Carnivora, Hyracoidea, Eulipotyphla, Pholidota, Primates, Proboscidea and Rodentia), with most of the detected taxa categorized within the orders of primates, carnivores, and rodents (Table1). From these, the order Hyracoidea was only detected in the bulk fly leg pool based destructive method and the orders Eulipotyphla, Pholidota and Proboscidea were in bulk fly body pool non-destructive method, and the rest were detected in both. In the prior investigation by Gogarten et al. (2020), a total of 69 mammal species were identified using a destructive method based on small fly pools (Entire fly pool destructive extraction method). After eliminating species overlap across all sampling sites, 48 distinct taxa remained. All identified taxa belonged to 19 taxonomic families of eight orders. From these orders, apart from Eulipotyphla, which was only detected in the present study, and Chiroptera that was only detected in the study by Gogarten et al. (2020), the remaining seven orders were detected in both studies. In the current study, from the 54 mammal species detected, 10 of them (*Cercopithecus mitis*, *Colobus guereza*, *Erythrocebus patas*, *Papio Anubis*, *Cercopithecus campbelli*, *Atherurus africanus*, *Chlorocebus sabaeus, Canis adustus, Phacochoerus.africanus, Cricetomys* sp.) were detected using both extraction methods. After excluding species overlap between the destructive and non-destructive methods and considering each species only once across the eight sampling sites, the bulk fly body pool non-destructive method yielded a higher count of mammal species (43) compared to the bulk fly

leg pool based destructive method (11). The bulk fly leg pool based destructive method typically detected a limited number of mammals at each site, ranging from 2 to 3. In contrast, the bulk fly body pool non-destructive approach consistently detects a broader range of mammal species, with counts ranging from 5 to 12 per site. In the previous study, the entire fly pool destructive extraction method employed revealed a similar range of mammal species per site, spanning from 5 to 15 (Gogarten et al. 2020). Only seven of the detected mammal species (*Cercopithecus mitis*, *Colobus guereza*, *Erythrocebus patas*, *Papio Anubis*, *Cercopithecus campbelli*, *Atherurus africanus*, *Chlorocebus sabaeus*) were detected both in the current study and in Gogarten et al (2020), likely reflecting the fact that the relatively small number of flies processed here have not fully characterized the diversity of mammals found at these sites.

The IUCN species status of the species detected revealed a similar pattern, with more threatened mammal species detected with the bulk fly body pool non-destructive method than the bulk fly leg pool based destructive method. Most of the mammal species detected with the bulk fly leg pool based destructive method belong to the least concerned category, except one critically endangered and one nearly threatened detection species, but this might simply reflect the very low number of detection events with this method. In contrast, the bulk fly body pool non-destructive method allowed for the detection of more threatened species classified as Nearly Threatened (n=5), Endangered (n=5), Critically Endangered (n=1), Vulnarable (n=4). Gogarten et al. (2020), detected a similar range of threatened mammal species with entire fly pool destructive method (Nearly Threatened, n= 8; Endangered, n=7; Critically Endangered, n=2; Vulnarable, n=3).

The identified mammals display a broad ecological spectrum, encompassing a variety of sizes from a larger species such as *Loxodonta africana* (African savannah elephant), to smaller ones like *Cricetomys emini* (Emin's pouched rat). Moreover, the detected mammals exhibit a range of dietary preferences including herbivores like *Loxodonta africana* (African savannah elephant), omnivores such as *Civettictis civetta* (African civet) or omnivorous frugivore like *Pan troglodytes* (Chimpanzee).

Species richness comparison between methods

We observed a significant difference between the number of species detected with each of the extraction methods (ANOVA; $F_{2,21}$ = 12.35; P < 0.001; Figure 1). The post hoc pairwise comparisons between the different levels of the extraction methods revealed significant differences between the Bulk fly leg pool based destructive and the other two methods (Bulk fly leg pool based destructive vs. bulk fly body pool non-destructive, P<0.01; Bulk fly leg pool based destructive vs. entire fly pool destructive extraction, P<0.001; Figure 1). We detected no difference between the entire fly pool destructive extraction and bulk fly body pool non-destructive method (P=0.760; Figure 1). Indeed, the identification rates of mammal species

were nearly equivalent for both the entire fly pool destructive extraction method ($N_{\text{species detections total}}$ = 67; $\bar{x}_{\text{species/site}}$ = 8.4; range = 5 to 15) and the bulk fly body pool non-destructive method ($N_{\text{species detections total}}$ =59; $\bar{x}_{\text{species/site}}$ =7.4; range = 1 to 12). On the contrary, we detected only 15 mammal species using the bulk fly leg pool based destructive method ($N_{\text{species detections total}}$ = 15; $\bar{x}_{\text{species/site}}$ =1.9; range = 1 to 3). Moreover, as the flies used were collected at the same location, all the mammal species (except *Dendrohyrax dorsalis*) detected in the bulk fly leg pool based destructive method were identical to the ones detected with the bulk fly body pool non-destructive protocol.

DISCUSSION

By employing two bulk fly iDNA extraction techniques coupled with a metabarcoding approach, we successfully identified a wide range of mammal species using a limited number of fly pools collected across eight locations in five sub-Saharan African countries. We then compared these results with the previously documented mammal communities obtained from the same locations (except one, which for logistical reasons was close but not identical) using iDNA obtained by destructively extracting DNA from smaller pools of flies (Gogarten et al. 2020). The successful application of these two previously unexplored extraction techniques for the analysis of fly iDNA to study mammal diversity contributes to the collective endeavor of enhancing the monitoring of terrestrial wildlife species with iDNA analyses (Ji et al. 2022; Srivathsan et al. 2022).

Our results show a considerable variation in the detection of mammal species across the different extraction methods. Between the two methods used in this study, the bulk fly body pool non-destructive method detected a higher number of mammal species compared to the bulk fly leg pool based destructive method. In contrast, when comparing it to the results from Gogarten et al. (2020) (i.e. entire fly pool destructive extraction method), the number of mammal detections using the non-destructive method is nearly identical. Likely, simply as a result of the lower mammal detection rate of the leg based approach, the bulk fly body pool non-destructive method demonstrated a greater capability in detecting mammal species listed as threatened by the IUCN. Our results suggests that mammal DNA is present in low amounts in fly legs, and that it is likely necessary to include the entire body of the fly during DNA extraction, to increase the likelihood of generating amplifiable mammal DNA. Consequently, of the two bulk extraction approaches explored here, we suggest that bulk fly body pool non-destructive method has the greater potential to make a meaningful contribution to conservation efforts.

In contrast, when comparing it to the results from Gogarten et al. (2020) (i.e. entire flies in small pools, destructively extracted), the number of mammal detections using the non-destructive method applied here is nearly identical. Although there were some differences in the species

detected at each site with these two methods (Table 1), our findings demonstrate that the bulk fly body non-destructive extraction method detected an equivalent diversity of mammal species as the approach that used destructive sampling of smaller pools of flies. This shows that pooling so many individual flies together for targeting iDNA does not influence negatively the mammal detection rates. This is seemingly in contrast to what has been shown using leeches (Schnell et al. 2012) and Rodgers et al. (2017), who had suggested that single fly metabarcoding could increase detection rates when compared to fly pooling-based methods; these discrepancies may in part reflect technical advances that have allowed for an increased number of replicates and deeper sequencing of larger pools. Similar to the findings in Gogarten et al. (2020) and numerous other studies using flies for mammal monitoring (Rodgers et al. 2017; Lee et al. 2015; Calvignac-Spencer et al. 2013), we successfully identified a diverse array of mammal categories, both large and small in body size, and spanning from terrestrial artiodactyls to arboreal primates, with the bulk extraction methods tested here, particularly the bulk fly body pool non-destructive method approach. Thus, our results suggest that depending on the needs of the study, bulk pooling of hundreds of flies might be an effective strategy for reducing costs. An added bonus of the non-destructive extraction approach, is that the fly pools remained intact, allowing for the subsequent reuse of the same samples. As a result, these non-destructive methods can be used effectively for museum specimens or with samples that are not easily available or rare to collect (Lynggaard et al. 2019; Gilbert et al. 2007).

The primary constraint in iDNA-based mammal and vertebrate species monitoring using carrion flies has been the expenses associated with sample processing. In general, iDNA based analyses using flies typically involve the collection of DNA from either single flies or small pool of flies (e.g. Rodgers et al. 2017; Lee et al. 2016; Schubert et al. 2015; but see Srivathsan et al. (2022) for an extraction free protocol). This necessitates significant use of consumables and time for the extraction and handling of samples, which, in turn, hinders the feasibility of employing fly iDNA for extensive-scale biomonitoring efforts. In Gogarten et al. (2020), 15 fly pools of 7 flies were used for each location, which was time consuming and costly. In contrast, in our study, for each method we used only a single pool per sampling site, aggregating the 105 flies into a single sample. These experiments were performed while maintaining strict contamination avoidance methods, including regular surface cleaning followed by UV light-based sterilization in between processing each sample for each locations. As a result, we managed to achieve a cost reduction of around 15 fold, as we required comparatively less time and consumables to extract the same number of flies. In addition, the time for laboratory processing of 840 flies divided in eight pools (i.e. one per site) was approximately 12 working days. To help others assess the feasibility of such experiments, this process involved multiple steps, which can be broken down roughly as follows: extraction of 8 pools / day, for a total of 2 days for total 16 pools; 2 step PCR followed by gel extractions and

purifications for 16 pools, with 4 replicates per pool, at a rate of 96 reactions / day = 4 days; library preparation for high-throughput sequencing for all positive replicates = 5 days, data analysis= 1 day. One of the major limitations of the tested bulk extraction protocols explored here, particularly the leg removal, is that they involved a lot of manual work with flies to separate the fly parts and that entailed more opportunities for cross contamination. Future studies using the whole body of the fly in large pools for the non-destructive DNA extraction method, will not need these steps, which will simplify the extraction step and further reduce the time required. It is worth noting that these steps were done manually, without the assistance of tools for automation; the ready availability of such tools will even further reduce the labor and time needed to complete these experiments. While there are certainly further technical improvements of these methods possible, largescale mammal terrestrial biomonitoring with iDNA is now clearly within reach and we echo calls that it is now time to begin exploring how these eDNA tools can be added to the monitoring toolkit of governments and conservationists (Lee et al. 2023; Stein et al. 2023).

Clearly, the application of bulk fly pooling based approaches show promise for mammal monitoring, but fly iDNA has also been used for pathogen detection at different scales (Mwakasungula et al. 2022; Hoffmann et al. 2017; Hoffmann et al. 2016; Knauf et al. 2016). Whether such bulk pooling would also be effective for the detection of pathogens represents an important area of future research. In this context, we recently also employed the bulk fly leg pooling scheme to screen primate pathogens in and around Kibale National Park, Uganda, with promising results (Jahan et al. 2022). One intriguing frontier for using fly iDNA for studies of pathogen ecology, involves testing the host range of a pathogen, by linking pathogen detection events within a fly with the mammalian detection events in the exact same flies (Jahan et al. 2023; Hoffmann et al. 2017); such information may be obscured when heavily pooling samples as done here, thus losing important data on potential host-pathogen associations. Depending on the specific aims of the study, individual fly extraction may end up being a more time intensive but still very useful approach.

Taken together, our results suggest that mass pooling of fly bodies and thereafter using a non-digestion buffer to extract DNA could simplify the traditional fly extraction method and therefore allowing for the adoption of fly iDNA for terrestrial biomonitoring at large scales for both wildlife biodiversity. The ongoing global decline in biodiversity represents a critical and existential challenge for humanity, ultimately undermining the stability of ecosystems and disrupting the services they provide. The adoption of simplified yet effective large-scale biodiversity monitoring approaches has the potential to exert a significant positive impact on conservation endeavors, by providing much needed high-resolution data on biodiversity. Methodological advances in the use of fly iDNA for terrestrial biomonitoring, like those presented here, have

the potential allow fly iDNA to be used for large scale biomonitoring across very different ecosystems.

REFERENCES

1. Abrams J F, Hörig L A, Brozovic R, Axtner J, Crampton-Platt A, Mohamed A, Wong S T, Sollmann R, Yu D W, Wilting A (2019): Shifting up a gear with iDNA: From mammal detection events to standardised surveys. J. Appl. Ecol., 56, 1637-1648. DOI: 10.1111/1365-2664.13411.
2. Andersen K, Bird K L, Rasmussen M, Haile J, Breuning-Madsen H, Kjaer K H, Orlando L, Gilbert M T, Willerslev E(2012): Meta-barcoding of 'dirt' DNA from soil reflects vertebrate biodiversity. Mol Ecol., 8, 1966-79. DOI: 10.1111/j.1365-294X.2011.05261.x.
3. Bohmann K, Evans A, Gilbert M T P, Carvalho G R, Creer S, Knapp M, Yu D W, de Bruyn M (2014): Environmental DNA for wildlife biology and biodiversity monitoring. Trends Ecol. Evol., 29, 358-367. DOI: 10.1016/j.tree.2014.04.003.
4. Bohmann K, Schnell I B, Gilbert M (2013): When bugs reveal biodiversity. Mol Ecol., 22(4), 909-911. DOI: 10.1111/mec.12221.
5. Boyer F, Mercier C, Bonin A, Le Bras Y, Taberlet P, Coissac E (2016): Obitools: a unix-inspired software package for DNA metabarcoding. Mol Ecol Resour., 16(1), 176-82. DOI: 10.1111/1755-0998.12428.
6. Butchart S H M, Walpole M, Collen B, Strien A V, Scharlemann J P W, Almond R E A, Baillie J E M, Bomhard B, Brown C, Bruno J, Carpenter K E, Carr G M, Chanson J, Chenery A M, Csirke J, Davidson N C, Dentener F, Foster M, Galli A, Galloway J N, Genovesi P, Gregory R D, Hockings M, Kapos V, Lamarque J F, Leverington F, Loh J, McGeoch M A, McRae L, Minasyan A, Morcillo M H, Oldfield T E E, Pauly D, Quader S, Revenga C, Sauer J R, Skolnik B, Spear D, Stanwell-Smith D, Stuart S N, Symes A, Tierney M, Tyrrell T D, Vié J C, Watson R (2010): Global biodiversity: Indicators of recent declines. Science, 328, 1164-1168. DOI:10.1126/science.1187512.
7. Calvignac-Spencer S, Leendertz F H, Gilbert M T P, Schubert G (2013a): An invertebrate stomach's view on vertebrate ecology: Certain invertebrates could be used as "vertebrate samplers" and deliver DNA-based information on many aspects of vertebrate ecology. BioEssays 35, 1004-1013. DOI: 10.1002/bies.201300060.
8. Calvignac-Spencer S, Merkel K, Kutzner N, Kühl H, Boesch C, Kappeler P M, Metzger S, Schubert G, Leendertz F H(2013b): Carrion fly-derived DNA as a tool for

comprehensive and cost-effective assessment of mammalian biodiversity. Mol Ecol., 22, 915-924. DOI: 10.1111/mec.12183.

10. Clare E L, Economou C K, Bennett F J, Dyer C E, Adams K, McRobie B, Drinkwater R, Littlefair J E (2022): Measuring biodiversity from DNA in the air. Curr Biol., 32(3), 693-700.e5. DOI: 10.1016/j.cub.2021.11.064.
11. Cutajar T P, Rowley J J L (2020): Surveying frogs from the bellies of their parasites: Invertebrate-derived DNA as a novel survey method for frogs. Global Ecology and Conservation, 22, e00978-e00978. DOI: 10.1016/j.gecco.2020.e00978.
12. Danabalan R, Merkel K, Bærholm Schnell I, Arandjelovic M, Boesch C, Brazzola G, Dieguez P, Dupain J, Kambale-Vyalengerera M, Kühl H S, Hoffmann C (2023): Mammal mitogenomics from invertebrate-derived DNA. Environmental DNA. DOI: 10.1002/edn3.436.
13. Gariepy T D, Lindsay R, Ogden N, Gregory T R (2012): Identifying the last supper: Utility of the DNA barcode library for bloodmeal identification in ticks. Molecular Ecology Resources, 12, 646-652. DOI: 10.1111/j.1755-0998.2012.03140.x.
14. Gilbert M T, Moore W, Melchior L, Worobey M (2007): DNA extraction from dry museum beetles without conferring external morphological damage. PLoS One. 2(3), e272. DOI: 10.1371/journal.pone.0000272.
15. Gogarten J F, Düx A, Mubemba B, Pléh K, Hoffmann C, Mielke A, Müller-Tiburtius J, Sachse A, Wittig R M, Calvignac-Spencer S, Leendertz F H (2019): Tropical rainforest flies carrying pathogens form stable associations with social nonhuman primates. Mol Ecol., 18, 4242-4258. DOI: 10.1111/mec.15145.
16. Gogarten J F, Hoffmann C, Arandjelovic M, Sachse A, Merkel K, Dieguez P, Agbor A, Angedakin S, Brazzola G, Jones S, Langergraber K E, Lee K, Marrocoli S, Murai M, Sommer V, Kühl H, Leendertz F H, Calvignac-Spencer S (2020): Fly-derived DNA and camera traps are complementary tools for assessing mammalian biodiversity. Environmental DNA, 2, 63-76. DOI: 10.1002/edn3.46
17. Goldberg C S, Pilliod D S, Arkle R S, Waits L P (2011): Molecular detection of vertebrates in stream water: a demonstration using Rocky Mountain tailed frogs and Idaho giant salamanders. PLoS One. 6(7), e22746. DOI: 10.1371/journal.pone.0022746.
18. Gollan J R, de Bruyn L L, Reid N, Wilkie L (2013): Monitoring the ecosystem service provided by dung beetles offers benefits over commonly used biodiversity metrics and

a traditional trapping method. Journal for Nature Conservation, 21, 183-188. DOI: 10.1016/j.jnc.2012.12.004.

19. Grubaugh N D, Sharma S, Krajacich B J, Iii L S F, Bolay F K, Ii J W D, Johnson W E, Ebel G D, Foy B D, Brackney D E (2015): Xenosurveillance : a novel mosquito-based approach for examining the human- pathogen landscape. PLoS Negl Trop Dis., 3, e0003628. DOI: 10.1371/journal.pntd.0003628.

20. Hagen M, Kissling W D, Rasmussen C, De Aguiar M A M, Brown L E, Carstensen D W, Alves-Dos-Santos I, Dupont Y L, Edwards F K, Genini J, Guimarães P R, Jenkins G B, Jordano P, Kaiser-Bunbury C N, Ledger M E, Maia K P, Marquitti F M D, McLaughlin Ó, Morellato L P C, O'Gorman E J, Trøjelsgaard K, Tylianakis J. M, Vidal M M, Woodward G, Olesen J M (2012): Biodiversity, Species Interactions and Ecological Networks in a Fragmented World. Adv. Ecol. Res, 46, 89-210. DOI: 10.1016/B978-0-12-396992-7.00002-2.

21. Hoffmann C, Merkel K, Sachse A, Rodríguez P, Leendertz F H, Calvignac-Spencer S (2018): Blow flies as urban wildlife sensors. Mol. Ecol. Resour., 18, 502-510. DOI: 10.1111/1755-0998.12754.

22. Hoffmann C, Stockhausen M, Merkel K, Calvignac-Spencer S, Leendertz F H (2016): Assessing the feasibility of fly based surveillance of wildlife infectious diseases. Scientific Reports, 6, 1-9. DOI: 10.1038/srep37952.

23. Hoffmann C, Zimmermann F, Biek R, Kuehl H, Nowak K, Mundry R, Agbor A, Angedakin S, Arandjelovic M, Blankenburg A, Brazolla G, Corogenes K, Couacy-Hymann E, Deschner T, Dieguez P, Dierks K, Düx A, Dupke S, Eshuis H, Formenty P, Yuh Y G, Goedmakers A, Gogarten J F, Granjon A C, McGraw S, Grunow R, Hart J, Jones S, Junker J, Kiang J, Langergraber K, Lapuente J, Lee K, Leendertz S A, Léguillon F, Leinert V, Löhrich T, Marrocoli S, Mätz-Rensing K, Meier A, Merkel K, Metzger S, Murai M, Niedorf S, De Nys H, Sachse A, van Schijndel J, Thiesen U, Ton E, Wu D, Wieler L H, Boesch C, Klee S R, Wittig R M, Calvignac-Spencer S, Leendertz F H (2017): Persistent anthrax as a major driver of wildlife mortality in a tropical rainforest.Nature, 548(7665), 82-86. DOI: 10.1038/nature23309.

24. Hothorn T, Bretz F, Westfall P (2008): Simultaneous inference in general parametric models. Biom J., 50(3), 346-363. DOI: 10.1002/bimj.200810425.

25. Jahan M, Calvignac-Spencer S, Chapman C A, Kalbitzer U, Leendertz F H, Omeja P A, Sarkar D, Ulrich M, Gogarten J F (2022): The movement of pathogen carrying flies at the human-wildlife interface. Ecohealth, 19(4), 450-457. DOI: 10.1007/s10393-022-01621-8.

26. Jahan M, Lagostina L, Gräßle T, Couacy-Hymann E, Kouadio L, Kouakou V K, Krou H A, Mossoun A M, Patrono L V, Pléh K and Steiner J A (2023): Fly iDNA suggests strict reliance of the causative agent of sylvatic anthrax on rainforest ecosystems. Environmental DNA, 1-12. DOI: 10.1002/edn3.401.

27. Ji Y, Baker C C M, Popescu V D, Wang J, Wu C, Wang Z, Li Y, Wang L, Hua C, Yang Z, Yang C, Xu C C Y, Diana A, Wen Q, Pierce N E, Yu D W (2022): Measuring protected-area effectiveness using vertebrate distributions from leech iDNA. Nat Commun., 13(1), 1555. DOI: 10.1038/s41467-022-28778-8.

28. Johnson M D, Katz A D, Davis M A, Tetzlaff S, Edlund D, Tomczyk S, Molano-Flores B, Wilder T, Sperry J H (2023): Environmental DNA metabarcoding from flowers reveals arthropod pollinators, plant pests, parasites, and potential predator–prey interactions while revealing more arthropod diversity than camera traps. Environmental DNA, 5, 551-569. DOI: 10.1002/edn3.411.

29. Kahle, D. and H. Wickham (2013): ggmap: Spatial visualization with ggplot2. R Journal, 5, 144-161. DOI: 10.32614/rj-2013-014.

30. Keesing F, Belden L K, Daszak P, Dobson A, Harvell C D, Holt R D, Hudson P, Jolles A, Jones K E, Mitchell C E, Myers S S, Bogich T, Ostfeld R S (2010): Impacts of biodiversity on the emergence and transmission of infectious diseases. Nature, 468(7324), 647-652. DOI: 10.1038/nature09575.

31. Knauf S, Raphael J, Mitjà O, Lejora I AV, Chuma I S, Batamuzi E K, Keyyu J D, Fyumagwa R, Lüert S, Godornes C, Liu H, Schwarz C, Šmajs D, Grange P, Zinner D, Roos C, Lukehart S A (2016): Isolation of treponema dna from necrophagous flies in a natural ecosystem. EBio Medicine, 85-90. DOI: 10.1016/j.ebiom.2016.07.033.

32. Kocher A, de Thoisy B, Catzeflis F, Valière S, Bañuls A L, Murienne J (2017): iDNA screening: Disease vectors as vertebrate samplers. Mol Ecol., 22, 6478-6486. DOI: 10.1111/mec.14362.

33. Lassen S B, Nielsen S A, Skovgård H, Kristensen M (2011): Molecular identification of bloodmeals from biting midges (Diptera: Ceratopogonidae: Culicoides Latreille) in Denmark. Parasitology Research, 108, 823-829. DOI: 10.1007/s00436-010-2123-4

34. Lee K N, Kelly R P, Demir-Hilton E, Laschever E, Allan E A (2023): Adoption of environmental DNA in public agency practice. Environmental DNA, 00, 1–14. https://doi.org/10.1002/edn3.470

35. Lee P S, Gan H M, Clements G R, Wilson J J (2016): Field calibration of blowfly-derived DNA against traditional methods for assessing mammal diversity in tropical forests.

Genome, 59(11),1008-1022. DOI: 10.1139/gen-2015-0193. Epub 2016 May 11. PMID: 27696907

36. Lee P S, Sing K W, Wilson J J (2015): Reading mammal diversity from flies: the persistence period of amplifiable mammal mtDNA in blowfly guts (*Chrysomya megacephala*) and a new DNA mini-barcode target. PLoS One.10(4), e0123871. DOI: 10.1371/journal.pone.0123871.

37. Leempoel K, Hebert T, Hadly E A (2020): A comparison of eDNA to camera trapping for assessment of terrestrial mammal diversity. Proceedings of the Royal Society B: Biological Sciences, 287. DOI: 10.1098/rspb.2019.2353.

38. Lyimo I N, Ferguson H M (2009): Ecological and evolutionary determinants of host species choice in mosquito vectors. Trends in Parasitology, 25, 189-196. DOI: 10.1016/j.pt.2009.01.005.

39. Lyman J A, Sanchez D E, Hershauer S N, Sobek C J, Chambers C L, Zahratka J, Walker F M (2022): Mammalian eDNA on herbaceous vegetation? Validating a qPCR assay for detection of an endangered rodent. Environmental DNA, 4, 1187-1197. DOI: 10.1002/edn3.331.

40. Lynggaard C, Bertelsen M F, Jensen C V, Johnson M S, Frøslev T G, Olsen M T, Bohmann K (2022a): Airborne environmental DNA for terrestrial vertebrate community monitoring. Curr Biol., 32(3), 701-707. DOI: 10.1016/j.cub.2021.12.014.

41. Lynggaard C, Calvignac-Spencer S, Chapman C A, Kalbitzer U, Leendertz F H, Omeja P A, Opito E A, Sarkar D, Bohmann K, Gogarten J F (2023): Vertebrate environmental DNA from leaf swabs. Curr Biol., 33(16), R853-R854. DOI: 10.1016/j.cub.2023.06.031.

42. Lynggaard C, Frøslev T G, Johnson M S, Olsen M T, Bohmann K (2022b): Airborne environmental DNA captures terrestrial vertebrate diversity in nature. Mol Ecol Resour. DOI: 10.1111/1755-0998.13840.

43. Lynggaard C, Nielsen M, Santos-Bay L, Gastauer M, Oliveira G, Bohmann K (2019): Vertebrate diversity revealed by metabarcoding of bulk arthropod samples from tropical forests. Environmental DNA, 1, 329-341. DOI: 10.1002/edn3.34.

44. Macher T H, Schütz R, Hörren T, Beermann A J, Leese F (2023): It's raining species: Rainwash eDNA metabarcoding as a minimally invasive method to assess tree canopy invertebrate diversity. Environmental DNA, 5, 3-11. DOI: 10.1002/edn3.372.

45. Massey A L, Bronzoni R V M, da Silva D J F, Allen J M, de Lázari P R, Dos Santos-Filho M, Canale G R, Bernardo C S S, Peres C A, Levi T (2022): Invertebrates for

vertebrate biodiversity monitoring: Comparisons using three insect taxa as iDNA samplers. Mol Ecol Resour., 22(3), 962-977. DOI: 10.1111/1755-0998.13525.

46. Mwakasungula S, Rougeron V, Arnathau C, Boundenga L, Miguel E, Boissière A, Jiolle D, Durand P, Msigwa A, Mswata S, Olotu A, Sterkers Y, Roche B, Killeen G, Cerqueira F, Bitome-Essono P Y, Bretagnolle F, Masanja H, Paupy C, Sumaye R, Prugnolle F (2022): Using haematophagous fly blood meals to study the diversity of blood-borne pathogens infecting wild mammals. Mol Ecol Resour., 22(8), 2915-2927. DOI: 10.1111/1755-0998.13670.

47. Newton J P, Bateman P W, Heydenrych M J, Kestel J H, Dixon K W, Prendergast K S, White N E, Nevill P (2023): Monitoring the birds and the bees: Environmental DNA metabarcoding of flowers detects plant–animal interactions. Environmental DNA, 5, 488-502. DOI: 10.1002/edn3.399.

48. Nielsen M, Gilbert M T P, Pape T, Bohmann K (2019): A simplified DNA extraction protocol for unsorted bulk arthropod samples that maintains exoskeletal integrity. Environmental DNA, 1, 144-154. DOI: 10.1002/edn3.16.

49. Owings C G, Banerjee A, Asher T M D, Gilhooly W P 3rd, Tuceryan A, Huffine M, Skaggs C L, Adebowale I M, Manicke N E, Picard C J (2019): Female blow flies as vertebrate resource indicators. Sci Rep., 9(1), 10594. DOI: 10.1038/s41598-019-46758-9.

50. R Core Team (2021): R: A language and environment for statistical computing. R Foundation for Statistical Computing. https://www.R- proje ct.org/.

51. Rees H C, Maddison B C, Middleditch D J, Patmore J R M, Gough K C (2014): The detection of aquatic animal species using environmental DNA- a review of eDNA as a survey tool in ecology. Journal of Applied Ecology, 51,1450-1459. DOI: 10.1111/1365-2664.12306.

52. Rodgers T W, Xu C C Y, Giacalone J, Kapheim K M, Saltonstall K, Vargas M, Yu D W, Somervuo P, McMillan W O, Jansen P A (2017): Carrion fly-derived DNA metabarcoding is an effective tool for mammal surveys: Evidence from a known tropical mammal community. Molecular Ecology Resources 17, 133-145. DOI: 10.1111/1755-0998.12701.

53. Schnell I B, Bohmann K, Schultze S E, Richter S R, Murray D C, Sinding M S, Bass D, Cadle J E, Campbell M J, Dolch R, Edwards D P, Gray T N E, Hansen T, Hoa A N Q, Noer C L, Heise-Pavlov S, Sander Pedersen A F, Ramamonjisoa J C, Siddall M E, Tilker A, Traeholt C, Wilkinson N, Woodcock P, Yu D W, Bertelsen M F, Bunce M, Gilbert M T P (2018): Debugging diversity- a pan-continental exploration of the potential

of terrestrial blood-feeding leeches as a vertebrate monitoring tool. Mol Ecol Resour., 18(6), 1282-1298. DOI: 10.1111/1755-0998.12912.

54. Schnell I B, Thomsen P F, Wilkinson N, Rasmussen M, Jensen L R D, Willerslev E, Bertelsen M. F, Gilbert M T P (2012): Screening mammal biodiversity using DNA from leeches. Current Biology 22, R262-R263. DOI: 10.1016/j.cub.2012.02.058.

55. Schubert G, Stockhausen M, Hoffmann C, Merkel K, Vigilant L, Leendertz F H, Calvignac-Spencer S (2015): Targeted detection of mammalian species using carrion fly-derived DNA. Mol Ecol Resour., 15(2), 285-94. DOI: 10.1111/1755-0998.12306.

56. Srivathsan A, Loh R K, Ong E J, Lee L, Ang Y, Kutty S N, Meier R (2022): Network analysis with either Illumina or MinION reveals that detecting vertebrate species requires metabarcoding of iDNA from a diverse fly community. Mol Ecol., 1-18. DOI: 10.1111/mec.16767.

57. Stein E D, Jerde C L, Allan E A, Sepulveda A J, Abbott C L, Baerwald M R, Darling J, Goodwin K D, Meyer R S, Timmers M A, Thielen P M (2023): Critical considerations for communicating environmental DNA science. Environmental DNA, 00, 1–12. https://doi.org/10.1002/edn3.472

58. Taberlet P, Bonin A, Zinger L, Coissac E (2018): Environmental DNA: For biodiversity research and monitoring. Oxford University Press.

59. Ushio M, Fukuda H, Inoue T, Makoto K, Kishida O, Sato K, Murata K, Nikaido M, Sado T, Sato Y, Takeshita M, Iwasaki W, Yamanaka H, Kondoh M, Miya M (2017): Environmental DNA enables detection of terrestrial mammals from forest pond water. Molecular Ecology Resources, 17, e63-e75. DOI: 10.1111/1755-0998.12690.

60. Valentini A, Taberlet P, Miaud C, Civade R, Herder J, Thomsen P F, Bellemain E, Besnard A, Coissac E, Boyer F, Gaboriaud C, Jean P, Poulet N, Roset N, Copp G H, Geniez P, Pont D, Argillier C, BauDOIn J-M, Peroux T, Crivelli A J, Olivier A, Acqueberge M, Le Brun M, Møller P R, Willerslev E, Dejean T (2016): Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. Mol. Ecol., 25, 929-942. https://DOI.org/10.1111/mec.13428.

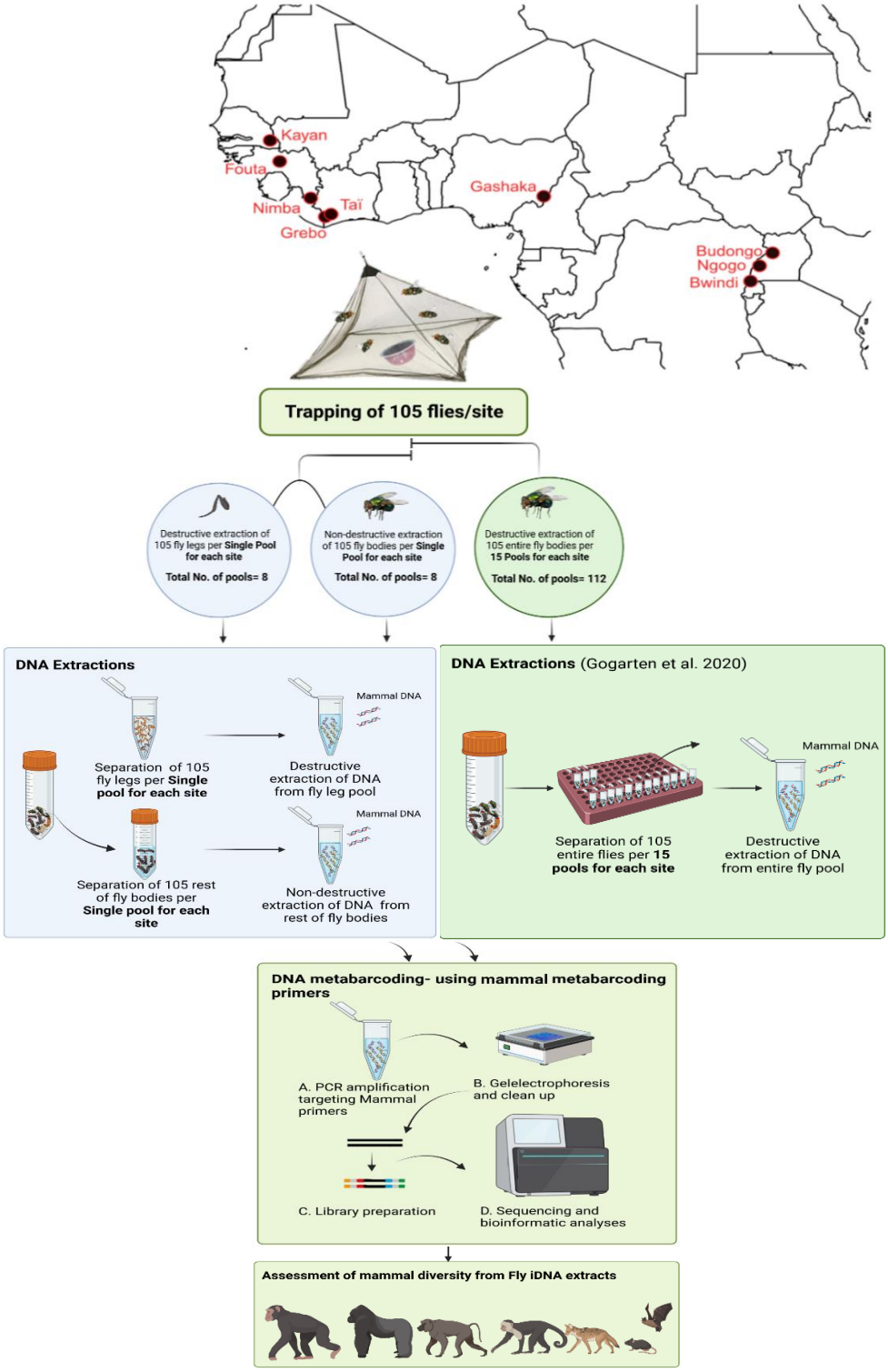

FIGURE 1: Experimental workflow along with the map of sampling sites. Blue represents the workflow of the present study in which two extraction methods were tested; 1. Bulk fly leg pool based destructive extraction and 2.Bulk fly body pool based non-destructive extraction. Green box represents the previous study showing entire fly based destructive extraction method.

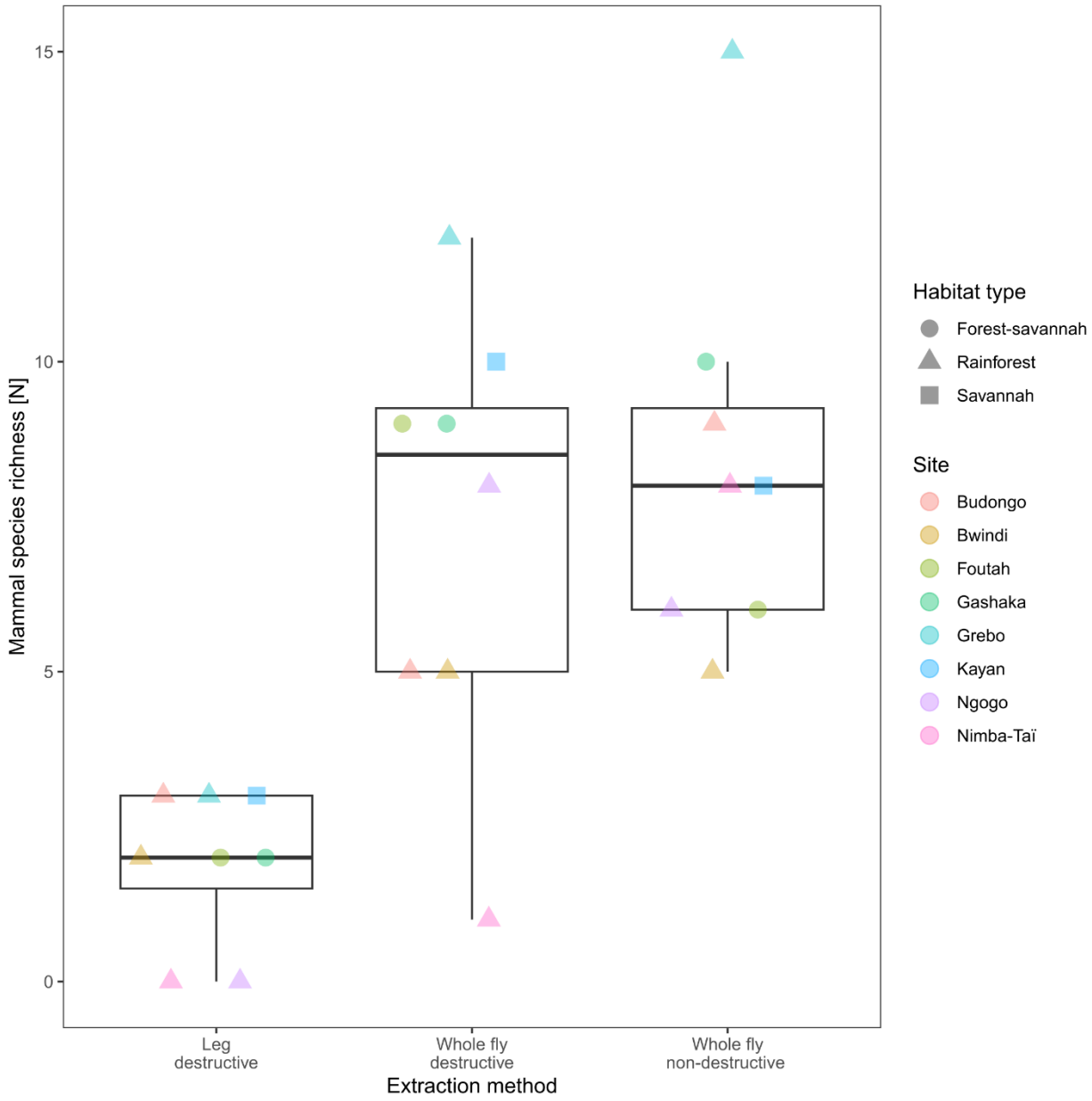

FIGURE 2: Box and whisker plot of mammal species richness per sampling site, separated by the different extraction methods used. The horizontal line in the whisker plots represents the median, while the lower and upper bounds of the boxes indicate the first and third quartiles, respectively. The upper and lower whiskers extend to the largest and lowest values no more than 1.5 times the interquartile range away from the hinge. Raw data are plotted as points, with the color and shape corresponding to the site and habitat type.

	Taxonomy of detected mammals					Method		
Location	Order	Family	Species	Common name	IUCN/GBIF status	Leg Destructive	Fly Non-destructive	Fly Destructive
Budongo, Uganda	Carnivora	Herpestidae	*Crossarchus alexandri*	Alexander's kusimanse	LC			✓
	Chiroptera	Molossidae	*Mops condylurus*	Angolan mops bat	LC			✓
		Pteropodidae	*Hypsignathus monstrosus*	Hammer-headed bat	LC			✓
	Hyracoidea	Procaviidae	*Dendrohyrax dorsalis*	Western tree hyrax	LC	✓		✓
	Primates	Cercopithecidae	*Cercopithecus albogularis*	Sykes monkey	UN		✓	✓
			Cercopithecus ascanius	Red tailed monkey	LC		✓	✓
			Cercopithecus mitis	Blue monkey	LC	✓	✓	✓
			Colobus guereza	Guereza monkey	LC	✓	✓	✓
		Hominidae	*Pan troglodytes*	Chimpanzee	EN		✓	✓
Bwindi, Uganda	Artiodactyla	Suidae	*Potamochoerus porcus*	Red river hog	LC			✓
	Carnivora	Canidae	*Canis adustus*	Side-striped jackal	LC		✓	
	Chiroptera	Molossidae	*Mops condylurus*	Angolan mops bat	LC			✓
	Primates	Cercopithecidae	*Cercopithecus mitis*	Blue monkey	LC	✓		
			Colobus guereza	Guereza monkey	LC	✓	✓	
		Hominidae	*Gorilla beringei*	Eastern mountain gorilla	CR			✓
			Pan troglodytes	Chimpanzee	EN			✓
	Proboscidea	Elephantidae	*Loxodonta africana*	African bush elephant	EN		✓	✓
	Rodentia	Muridae	*Praomys jacksoni*	Jackson's soft-furred mouse	LC		✓	
		Sciuridae	*Heliosciurus ruwenzorii*	Ruwenzori sun squirrel	LC		✓	
Fouta, Uganda	Carnivora	Canidae	*Canis adustus*	Side-striped jackal	LC		✓	✓
		Herpestidae	*Herpestes ichneumon*	Egyptian mongoose	LC		✓	
			Mungos mungo	Banded mongoose	LC		✓	
	Eulipotyphla	Soricidae	*Crocidura lamottei*	Lamotte's shrew	LC		✓	
	Hyracoidea	Procaviidae	*Procavia capensis*	Rock hyrax	LC			✓
	Primates	Cercopithecidae	*Chlorocebus sabaeus*	Green monkey	LC		✓	✓
			Erythrocebus patas	Southern patas monkey	CR	✓	✓	✓
			Papio papio	Guniea baboon	NT		✓	
		Galagidae	*Galago senegalensis*	Bush baby	LC		✓	
	Rodentia	Hystricidae	*Atherurus africanus*	African brush-tailed porcupine	LC	✓		
		Muridae	*Mastomys natalensis*	Natal multimammate mouse	LC		✓	
		Nesomyidae	*Cricetomys emini*	Emin's pouched rat	LC			✓
		Sciuridae	*Heliosciurus rufobrachium*	Squirrel	LC			✓
Gashaka, Nigeria	Artiodactyla	Bovidae	*Cephalophus silvicultor*	Yellow-backed duiker	NT			✓
			Kobus ellipsiprymnus	Water buck	LC		✓	
			Syncerus caffer	African buffalo	NT		✓	
			Tragelaphus scriptus	Bush buck	LC		✓	✓
			Potamochoerus porcus	Red river hog	LC			✓
	Carnivora	Herpestidae	*Herpestes ichneumon*	Egyptian mongoose	LC			✓
		Viverridae	*Civettictis civetta*	African civet	LC			✓
	Chiroptera	Molossidae	*Mops condylurus*	Angolan mops bat	LC			✓
		Pteropodidae	*Hypsignathus monstrosus*	Hammer-headed bat	LC			✓
	Primates	Cercopithecidae	*Cercopithecus mona*	Mona monkey	NT			✓
			Cercopithecus nictitans	Putty nosed monkey	NT		✓	✓
			Papio anubis	Olive baboon	LC	✓	✓	✓
			Procolobus verus	Olive colobus	VU		✓	
		Lorisidae	*Perodicticus potto*	West african potto	NT		✓	
	Rodentia	Anomaluridae	*Anomalurus derbianus*	Lord derby's scaly-tailed squirrel	LC		✓	
		Hystricidae	*Atherurus africanus*	African brush-tailed porcupine	LC	✓	✓	

	Taxonomy of detected mammals					Method		
Location	**Order**	**Family**	**Species**	**Common name**	**IUCN/GBIF status**	**Leg Destructive**	**Fly Non-destructive**	**Fly Destructive**
Grebo, Liberia	Artiodactyla	Bovidae	*Cephalophus dorsalis*	Bay duiker	NT			✓
			Philantomba maxwellii	Maxwell's duiker	LC		✓	✓
			Syncerus caffer	African buffalo	NT		✓	
		Suidae	*Potamochoerus porcus*	Red river hog	LC			✓
	Carnivora	Herpestidae	*Crossarchus obscurus*	Common kusimanse	LC			✓
		Nandiniidae	*Nandinia binotata*	African palm civet	LC			✓
		Viverridae	*Genetta pardina*	Pardine genet	LC			✓
	Pholidota	Manidae	*Phataginus tricuspis*	White bellied pangolin	EN		✓	
	Primates	Cercopithecidae	*Cercocebus atys*	Sooty mangabey	VU		✓	✓
			Cercopithecus campbelli	Campbell's monkey	NT	✓	✓	✓
			Cercopithecus diana	Diana monkey	EN			✓
			Cercopithecus petaurista	Lesser spot-nosed monkey	NT			✓
			Piliocolobus badius	Western red colobus	EN		✓	✓
			Procolobus verus	Olive colobus	VU		✓	✓
		Lorisidae	*Perodicticus potto*	West african potto	NT		✓	✓
	Rodentia	Anomaluridae	*Anomalurus derbianus*	Lord derby's scaly-tailed squirrel	LC		✓	
		Hystricidae	*Atherurus africanus*	African brush-tailed porcupine	LC	✓	✓	✓
		Nesomyidae	*Cricetomys sp.*	African savannah pouched rat	LC	✓	✓	
		Sciuridae	*Protoxerus stangeri*	Forest giant squirrel	LC		✓	✓
Kayan, Senegal	Artiodactyla	Bovidae	*Hippotragus equinus*	Roan antelope	LC		✓	✓
			Sylvicapra grimmia	Common duikar	LC		✓	
			Tragelaphus scriptus	Bush buck	LC			✓
		Suidae	*Phacochoerus africanus*	Common warthog	LC	✓	✓	
	Carnivora	Canidae	*Canis adustus*	Side-striped jackal	LC	✓	✓	
		Herpestidae	*Mungos mungo*	Banded mongoose	LC		✓	
		Viverridae	*Civettictis civetta*	African civet	LC		✓	✓
			Genetta thierryi	Hausa genet	LC			✓
	Primates	Cercopithecidae	*Cercopithecus campbelli*	Campbell's monkey	NT			✓
			Chlorocebus sabaeus	Green monkey	LC	✓	✓	✓
			Papio papio	Guniea baboon	NT		✓	
		Galagidae	*Galago senegalensis*	Bush baby	LC			✓
	Rodentia	Nesomyidae	*Cricetomys emini*	Emin's pouched rat	LC		✓	
		Sciuridae	*Xerus erythropus*	Striped ground squirrel	LC			✓
		Thryoxmyidae	*Thryoxmys swinderianus*	Greater cane rat	LC		✓	
Ngogo, Uganda	Carnivora	Felidae	*Leptailurus serval*	Serval cat	LC		✓	
	Chiroptera	Pteropodidae	*Eidolon helvum*	Straw-coloured fruit bat	NT			✓
	Primates	Cercopithecidae	*Allochrocebus lhoesti*	L'hoest's monkey	VU		✓	✓
			Cercopithecus albogularis	Sykes monkey	UN		✓	✓
			Cercopithecus ascanius	Red-tailed monkey	LC		✓	✓
			Colobus guereza	Guereza monkey	LC		✓	✓
			Papio anubis	Olive baboon	LC		✓	
		Hominidae	*Pan troglodytes*	Chimpanzee	EN			✓
	Proboscidea	Elephantidae	*Loxodonta africana*	African bush elephant	EN		✓	
	Rodentia	Muridae	*Praomys jacksoni*	Jackson's soft-furred mouse	LC		✓	
Nimba-Taï, Liberia-Ivory Coast	Artiodactyla	Bovidae	*Cephalophus jentinki*	Jentink's duiker	EN		✓	
	Carnivora	Herpestidae	*Crossarchus obscurus*	Common kusimanse	LC			✓
		Herpestidae	*Herpestes ichneumon*	Egyptian mongoose	LC			✓
		Nandiniidae	*Nandinia binotata*	African palm civet	LC			✓
	Pholidota	Manidae	*Phataginus tricuspis*	white bellied pangolin	EN			✓
	Primates	Galagidae	*Galagoides demidovii*	Demidoff's bushbaby	LC			✓
		Lorisidae	*Perodicticus potto*	West african potto	NT			✓
	Rodentia	Hystricidae	*Atherurus africanus*	African brush-tailed porcupine	LC			✓
		Sciuridae	*Protoxerus stangeri*	Forest giant squirrel	LC			✓

TABLE1: Mammal species detected at eight sites of sub-Saharan Africa using fly iDNA metabarcoding. The table indicates which of the extraction methods the particular species was detected with at each site. Information on the IUCN red list status of these species is provided: LC (least concern), EN (endangered), VU (vulnerable), CR (critically endangered), NT (near threatened) and UN (Unknown) are shown here.

4 Discussion

Monitoring wild animal populations and their pathogens is critical for conservation and human health. Here, I have begun exploring how molecular tools, in particular the use of environmental DNA and more specifically iDNA, can help us understand animals and their pathogens in a rapidly changing world. The need for wildlife and pathogen monitoring is particularly acute in the understudied and highly biodiverse tropics, areas where resources are often inadequate and animals are often more elusive because of their experience with hunting. As such, an important aspect of this thesis was to explore how these tools can be further optimized to allow them to be more readily deployed in resource poor regions.

This dissertation has investigated the potential utility of fly iDNA for investigating the disease ecology and distribution of a specific pathogen, namely Bcbva, in the TNP ecosystem (Chapter 1), while also exploring the potential of flies to serve as a vector, transmitting primate pathogens to human use areas surrounding the Kibale National Park in Uganda (Chapter 2). Lastly, this thesis focused on optimizing and improving fly-based biomonitoring systems, both to bring down costs by maximizing the number of flies that could be extracted in a single experiment, while also exploring non-destructive sampling that might open up museum specimens to explore changes in biodiversity through time (Chapter 3). For this purpose, I employed two new extraction methods of flies, and tested their efficacy for reconstructing mammal diversity in different tropical ecosystems in Sub-Saharan Africa.

These projects further demonstrate the power of fly iDNA based biomonitoring to effectively assess the mammal species diversity in different ecosystems, while showcasing how these types of data can be integrated with pathogen monitoring to provide insights into disease ecology and our understanding of human disease risk. The three studies that were the focus of my thesis helped further establish fly iDNA as a tool for mammal and pathogen surveillance along gradients of anthropogenic disturbance and raises the important point to continue explore the potential role of flies in disease transmitting pathogens, given that viable pathogens are often found in flies.

4.1 Small scale in-depth investigation of Bcbva ecology with fly iDNA

Metabarcoding of fly iDNA has been increasingly used as a tool for exploring mammal diversity, and for detecting pathogens as well. Taking both of these sources of information into consideration, I explored the potential for fly iDNA to contribute to investigating a novel or emerging infectious pathogen's ecology and distribution in the first chapter of this thesis. Humans and livestock living near TNP have been shown to be seropositive for a pathogen that has caused massive mortality in wildlife, Bcbva (Dupke et al. 2020). Little is known about the circulation of Bcbva outside of the protected national park. In addition, viable anthrax has been detected in roughly 5% of the flies inside TNP, suggesting flies might be a route of exposure for humans living around the park as well. To understand where human exposure might occur, I used fly iDNA to study the distribution of Bcbva along a gradient spanning from the pristine forest within Taï National Park, Côte d'Ivoire, out to surrounding villages. I also examined how iDNA can enhance genomic surveillance of Bcbva and explored whether fly iDNA can recapitulate our knowledge about the host range of this pathogen. Typically, sampling carcasses for this pathogen is challenging, as Bcbva is a BSL3 level pathogen and an ebola virus has also caused wildlife mortality in this park, meaning that necropsies need to be performed under very high biosafety standards. The use of fly iDNA thus represents a much faster and lower risk procedure that can be performed more readily by non-experts, an important consideration for scaling up monitoring efforts. This work involved generating complete genomes of the identified isolates and integrating this diversity into known Bcbva genetic diversity, showcasing how these iDNA samples can scale up pathogen surveillance.

For this work, around 2,350 flies were trapped from nine trapping locations along three parallel transects covering three different habitats, on alternate days from the end of July to the beginning of September 2021, during the rainy season. These samples could be easily collected by teams conducting other field work in the area, showcasing the potential for fly iDNA to contribute to biomonitoring. Specifically, three trapping areas were selected in villages' periphery, with traps located ~100 m away from the last house of the three villages (hereafter referred to as village habitat). Such areas can be difficult to sample and fly collection represents a very straightforward and relatively non-invasive sampling procedure. Using fly iDNA, we were thus able to generate

insights into habitats at the park edge and around villages that have been historically understudied in terms of Bcbva prevalence and the existing mammal diversity.

Using fly iDNA metabarcoding, I was able to reconstruct the mammal and fly species communities from the same fly pools with the aim of exploring possible links with Bcbva detection across three habitats. The Bcbva detected through iDNA spanned a substantial portion of the pathogen's known genomic diversity, which has been generated with decades monitoring carcasses (Hoffmann et al. 2017). Likewise, the mammal hosts identified in Bcbva-positive fly pools exhibited significant similarity with the mammal species whose carcasses contained Bcbva (Hoffmann et al. 2017). This represents a relatively underutilized approach for iDNA, and the fact that we largely recapitulated the findings of over 20 years of carcass monitoring in the region, highlights the power of iDNA for contributing to our understanding of disease ecology for less well studied pathogens as well.

Interestingly, in this study I also found that the detection rate of Bcbva varied across the habitat types and a similar pattern was observed for the mammal and fly diversity. This provides insights both into the areas where Bcbva is likely active, while also shedding light onto areas where humans are likely getting exposed to Bcbva, as suggested by serological studies (Dupke et al. 2020). Indeed, this study suggests that Bcbva may be largely restricted to the park and that human exposure is likely occurring in the forested area.

Perhaps unsurprisingly, the forest habitat where the Bcbva detection rate was highest, also harboured the highest mammal detection rate and mammal species richness. Despite being intuitive, showing such patterns represents a critical dilemma for conservation efforts; rather than a dilution of this infectious diseases in the protected area, as has been suggested by the dilution hypothesis (Civitello et al. 2015), our study suggests that the park represents a hotspot for Bcbva. This suggests that the protected area may harbour more wildlife, which in turn can harbour Bcbva. These findings are in many ways similar to a large scale fly iDNA based monitoring study performed within TNP, which suggested a higher Bcbva positivity rate in the research area where hunting is rare (Hoffmann et al. 2017). Both of these studies suggest that higher mammal species diversity and mammal biomass in the protected area, may contribute to increased Bcbva mortality within this habitat. The potential of protected areas to serve as reservoirs for infectious disease represents an important finding and suggests that conservation efforts should incorporate disease mitigation efforts into their management plans, in an effort to offset any potential disease risk posed by conservation.

While our results on mammal biodiversity and anthrax prevalence are intuitive and recapitulate larger caracass and biodiversity monitoring efforts in the region, it is important to note potential limitations to fly-based iDNA disease ecology studies. One is that flies are mobile (explored in chapter 2) and flies can certainly move further than the gradient studied here. At the same time, the presence of animal DNA does not necessarily indicate the animal was alive in that habitat and it is possible that DNA comes from animals hunted in the forest but processed in the village. Detecting non-human primate DNA in village flies could be a reflection of the consistent consumption of these species in the area (Refisch and Koné 2005). Indeed, this represents an exciting facet of this research; using iDNA to study human exposure to wildlife, by understanding the animals, either alive or dead, that humans coming into contact with.

In contrast to the patterns observed for mammal diversity, for flies we found that the forest edge exhibited higher diversity of fly MOTUs than both the forest and the village. While a number of factors might be driving this observed pattern, one interesting theory is that habitat heterogeneity at the forest edge may allow both forest and anthropogenic flies to thrive which also supports edge biased distribution of insects concept (Nguyen and Nansen 2018) Future explorations involving more data on the flies themselves will certainly be interesting in this regard. In addition, future studies are needed to understand whether particular fly species are more likely to carry Bcbva and whether the change in the fly community in these habitats might be responsible for the reduced Bcbva detection rates in the village habitats.

Overall, the findings in this chapter clearly showcase the effectiveness of iDNA in biomonitoring and for pathogen surveillance across gradients of anthropogenic disturbance. Coupled with genomic sequencing of the pathogen, fly iDNA could reveal a broad extent of the phylogenetic diversity of this pathogen that circulates in TNP, while fly iDNA illuminated the geographical extent of this pathogen, as well as host range of Bcbva, seeming to highlight the bacterium's strong dependence on the rainforest ecosystem.

4.2 Fly as a potential vector for pathogen transmission

In the second study, given that flies seem to be carrying viable Bcbva and have been implicated in the transmission of a diversity of pathogens, I sought out investigate the potential role of flies in transmitting emerging pathogens at human-wildlife interfaces. Previous studies have

suggested that flies are found at high densities in primate social groups (Gogarten et al. 2022; Gogarten et al. 2019) and carry viable Bcbva and at least detectible *Treponema pallidume pertenue* DNA. Furthermore, these studies suggested that flies form long-term associations with primate social groups, remaining with mangabeys for up to two weeks as the animals moved through the forest. This study focused on exploring the potential movement of nonhuman primate associated flies away from their social groups and into the landscapes around Kibale national Park, Uganda. Through a mark-recapture experiment, flies were marked with glo-germ powder in groups different nonhuman primate species at the forest edge and then we attempted to recapture them in more human intensified areas at distances of 50 m, 200 m, and 500 m from the marking sites. Surprisingly, given the proportional associations of these flies with nonhuman primates, we found that flies that marked in nonhuman primate groups were recaptured in some human use areas (19 of 28,615 recapture effort were marked flies).

The recapture rate of marked flies correspondent roughly with the distance from the marking sites. The highest number of marked flies were recaptured within 50m, while only 2 were observed at a 200m distance from where flies had been marked, and none were recaptured from 500m distance. This recapture rate was much lower than the fly recapture rate observed in previous mark-recapture experiment in a group of mangabey in TNP (Gogarten et al. 2019). Taken together, these findings seem to suggest that flies do show a preference for consistently engaging with social groups of nonhuman primates, but that they do occasionally leave these association and can move into anthropized areas.

The recapture rates of flies markedly decreased with time since the flies were marked. The flies were recaptured 2 and 4 days after initial marking, while but none were recaptured after 7,14 or 21 days, indicating either the diffusion of flies, the loss of marking powder through time or relatively short fly survival post-capture (due to the combination of short lifespans and age at capture, capture stress or exposure to the powder). Follow up experiments in the field under controlled conditions might be helpful to explore the impact of fly powder on fly survival, as well as the typical lifespan of these fly species. Such experiments were not performed here, in part due to the challenges of the SARS-CoV-2 pandemic. In addition, it might be that different flies associate with primate groups to varying degrees; here too additional studies collaborating with exerts in fly identification in the field, so that species can be marked separately, may be illuminating. A clear limitation of this study was that I was not able to perform molecular analyses on the particular fly species that were recaptured, as these were not collected in the field this precluded a determination of the specific flies which left the forest region or the parasites that those particular

flies carried. Future research holds the potential to provide an in depth understanding of the individual fly species carrying particular pathogens and these species' particular behavior and interactions with primates, information that the current study could not provide.

In addition, using metabarcoding of fly iDNA I detected the DNA of multiple eukaryotic primate parasites in the flies captured in nonhuman primate groups. In particular, three families of eukaryotic parasites (Blastocystidae, Entamoebidae and Vahlkampfiida) were detected from the flies captured in primate groups, while one additional parasite family (Trypanosomatidae) was detected outside the primate groups but within the forest area. The finding of pathogens in flies that infect the primates they associate with also echoes previous studies where sooty mangabey associated flies contained the DNA of *Treponema pallidum pertenue DNA* and vialble Bcbva (Gogarten et al. 2019), clearly these primate-associated flies come into contact with a diversity of pathogens during their lifecycles and there is the potential for transmission that requires further study. The fact that these primate-associated flies sometimes leave a primate social group and cross the human-wildlife interface into anthropogenic use areas, suggests that flies may play a role in transmitting pathogens across these interfaces. Despite this potential risk, it is important to note that we did not test for the infectivity of the pathogens whose DNA was detect on these flies, and this represents an important area of future research. The detection of parasite DNA does not prove infectivity of these parasites and future studies are needed to explore the actual disease risk posed by the pathogens found in and on nonhuman primate-associated flies at the human-wildlife interface.

In conclusion, these results demonstrate the potential of flies to serve as mechanical vector for emerging infectious agents at this human-wildlife interface and suggests further consideration as vectors between human and wildlife populations are warranted.

4.3 Optimizing fly extraction to scale up sample throughput

Based on the successful bulk fly extraction methods employed in Chapter 2 to explore fly associated disease risk, as well as the potential of fly iDNA based biomonitoring monitoring more broadly, in Chapter 3, I explored new avenues to optimize the fly iDNA based monitoring by scaling up the sampling throughput in a time and cost effective way. To this end, I explored possible ways of streamlining the fly iDNA based biodiversity monitoring techniques, emphasizing

different extraction approaches. I applied two distinct fly extraction techniques; bulk pooling of fly legs that could then be destructively extracted and bulk pooling of fly bodies that were then non-destructively extracted. These extracts were then examined using our typical metabarcoding approach, and compared to a previous study that examined flies from these ecosystems using a lower throughput extraction method (Gogarten et al. 2020). Both bulk approaches allowed me to identify a broad array of mammal species across these eight different locations in five sub-Saharan Africa, though the non-destructive whole body based approach appeared more promising and revealed a similar diversity as the lower throughput approach.

Metabarcoding of fly iDNA has already proven its potential in the field of terrestrial mammal (Rodgers et al. 2017; Calvignac-Spencer et al. 2013b) and pathogen diversity monitoring across different ecosystems (Gogarten et al. 2020; Kocher et al. 2017) and along gradients between habitats (Jahan et al. 2023; Ji et al. 2022). In general, iDNA based analyses in flies typically involve the collection of DNA from either single flies or small pool of flies (Rodgers et al. 2017; Lee et al. 2016; Schubert et al. 2015). This approach necessitates significant use of consumables and time for the extraction and handling of samples, which, in turn, hinders the feasibility of employing fly iDNA for extensive-scale biomonitoring efforts. In an effort to streamline the existing fly analysis methods, Srivathsan et al. (2022) proposed a much simplified, yet cost effective, extraction free protocol for analyzing individual live fly excreta to monitor vertebrate diversity. In another study, (Jangra and Ghosh 2022) demonstrated a cost free extraction protocol from individual soft bodied insects, which could facilitate fly iDNA based monitoring as well. Ji et al. (2022) employed non-destructive extraction of on average 34 leeches per pool, for a total of 893 replicates to describe the vertebral distribution across the entire collection in Ailaoshan reserve in Yunnan, China. In this context, the bulk fly pooling at a broad geographic range explored in Chapter 3 adds to this effort of simplifying fly iDNA studies and enabling intensified sampling throughput.

I analyzed a relatively small number of flies from each site with the abovementioned bulk extraction methods. A wide range of the mammal community was detected by both of bulk pool destructive and non-destructive methods. However, a greater number of mammal taxa were detected with the non-destructive sampling of fly body pools compared to the leg-based destructive protocol. In a prior study by Gogarten et al. (2020) a diverse range of mammal taxa was detected with an entire fly based destructive extraction method and smaller pooling scheme, from the same locations. It has been suggested that individual fly metabarcoding could increase detection rates when compared to fly pooling based methods (Rodgers et al. 2017); thus reducing

the size of the fly pools leads to higher mammal detection rates, but it also comes with increased costs. Our findings demonstrate that the non-destructive method detected nearly equal numbers of mammal species. Furthermore, when employing a non-destructive extraction approach, the fly pools remained intact, allowing for the subsequent reuse of the same samples with a very lower cost and time. As a result, these non-destructive methods can be used effectively for ancient samples or with samples that are not easily available or rare to collect (Lynggaard et al. 2019; Gilbert et al. 2007).

Similar to the findings in Gogarten et al. (2020) and many other fly iDNA based studies for mammal monitoring (Rodgers et al. 2017; Lee et al. 2015; Calvignac-Spencer et al. 2013b), we successfully identified a diverse array of mammal categories, spanning from terrestrial artiodactyls to arboreal primates, based on their habitat preferences. Moreover, we detected the same number of taxonomic orders except for one notable difference: with the non-destructive extraction method, we were able to detect the order Eulipotyphla, while the previously detected order Chiroptera was not picked up by the current methods; we strongly suspect that these differences are simply stochastic and do not represent a differential ability of the different extraction methods for detecting smaller animal's DNA. Indeed, both of the tested extraction techniques identified a substantial number of smaller mammals.

A major constraint in scaling up iDNA-based mammal and vertebrate species monitoring using carrion flies has been the expenses associated with sample processing. In Gogarten et al. (2020), 15 fly pools of 7 flies were used for each location and a total 112 fly pools were analysed for the study, which was relatively time consuming and costly. In contrast, in current study I used one single pool per sampling site by aggregating the 105 flies into a single pool for each method; thus only 16 pools in total were used for both the extraction methods. As a result, across the experiment, we managed to achieve a cost reduction of around 15 times, as we required comparatively less time and consumables to extract the same number of flies (Gogarten et al. 2020). In addition, the time used for laboratory processing of 840 flies divided in eight pools (i.e. one per site) for each extraction method, for a total of 16 pools for both methods combined, was approximately 12 working days. Both of the bulk pooling methods used in this study offer potential as cost-effective, quicker, and more convenient tools for detecting animal diversity across a broad geographical distribution.

These experiments were performed while maintaining strict contamination avoidance methods, including regular surface cleaning followed by UV light based sterilization in between processing each samples for each locations. One of the major limitations of the fly leg bulk extraction

protocols involves lot of manual work during the separation of fly parts, which increases risk of cross contamination. More broadly, eDNA approaches are detecting trace amounts of DNA and this means that they are very sensitive to contamination; hence, both stringent protocols, laboratory infrastructure separating components of these experiments, and stringent cleaning protocols were implemented (Taberlet et al. 2018). These are challenging to maintain in countries with limited resources and finding ways to reproduce these experiments in Bangladesh is an area of work I am very excited to pursue in the future.

As shown in my other thesis chapters, fly iDNA based monitoring is a really promising tool for disease monitoring. Fly iDNA has proved to be effective for the detection and monitoring of wildlife pathogens including novel infectious pathogen like *Bacillus cereus* biovar *anthracis* (Bcbva) (Hoffmann et al. 2017). Besides detecting host pathogen, fly iDNA can also help in exploring the ecology and epidemiology of mammal pathogens in different ecosystems (Hoffmann et al. 2017; Hoffmann et al. 2016), particularly when linking host range with pathogen detection events. This sort of information may be obscured when heavily pooling samples as done here. Depending on the specific aims of the study, individual fly extraction may end up being a time intensive but useful method.

Overall, global biodiversity declines represent an existential crisis that places the survival of wildlife at risk and threatens the stability of tropical biodiversity hotspots. Implementing simplified but effective large-scale biodiversity monitoring methods can significantly influence conservation efforts. Taken together, our results suggest that mass pooling of fly bodies and thereafter using a non-digestion buffer to extract DNA could simplify the traditional fly extraction method and therefore allowing for the adoption of fly iDNA for terrestrial biomonitoring at large scales for both wildlife and pathogen diversity. Bulk sample pooling represents a promising tool for streamlining the assessment of wildlife and pathogen distribution across vast geographic regions.

5 Summary

Wildlife and disease monitoring plays a vital role in preparing and understanding our changing world; not only for gaining insights into ecological dynamics and pandemic preparedness, but also for formulating effective strategies to combat the global decline in biodiversity, particularly in resource poor tropical biodiversity hotspots. Invertebrate sources of eDNA, such as fly iDNA have demonstrated their value as a cost and resource efficient biomonitoring tool. They are not only effective in tracking wildlife diversity and identifying wild pathogens, but also hold promise for in-depth exploration of various aspects of wildlife and pathogen ecology. Considering the successful implementation of fly iDNA in previous studies for both of these aspects animal and pathogen detection in isolation, here I examined if fly iDNA can provide insights into disease ecology along different habitat types. I also explored the potential for flies to serve as a mechanical vector along these different habitats, in particular looking at primate associated flies, and their potential to move pathogens from the pristine forest to human use areas. Lastly, I explored potential ways to optimize existing fly iDNA analysis schemes, so that they can maximizing the sampling throughput and reduce costs, in an effort to enable broader-scale biodiversity assessments.

For the first study, fly iDNA were molecularly analysed to investigate the geographic distribution of a sylvatic anthrax causing pathogen, *Bacillus cereus* biovar *anthracis* (Bcbva) and revealed significant variation in the detection of Bcbva across the habitats spanning from Taï rainforest to the nearby villages. In addition, whole genome sequencing of Bcbva from positive fly pools revealed considerable overlap with the genomic diversity of Bcbva described with decades of surveillance efforts. Metabarcoding of fly iDNA was also conducted to analyse the composition of both fly and mammal communities along this ecological gradient, with the goal of exploring potential associations with the detection of Bcbva. The highest detection of Bcbva was found in the area where the most diverse mammal community was detected, in the forest. No Bcbva was detected in the village habitat, where the least diverse mammal community was detected, suggesting a strong reliance of Bcbva on the rainforest ecosystem. However, the study also demonstrated the highest fly community at the edge of the forest, with flies from both anthropogenic and the forest areas found at this interface, which highlighted the potential risk of flies to act as disease transmitting vector between rainforest wildlife and the human habitants in surrounding villages.

Therefore, in our second study, I conducted a mark-recapture experiment to study the movement patterns of flies associated with non-human primates in the vicinity of Kibale National Park, Uganda. Flies Marked within non-human primate groups were subsequently recaptured in areas used by humans, with 19 marked flies were captured with a recapture effort of 28,615 recapture effort flies. The application of metabarcoding on the flies collected within non-human primate groups unveiled the presence of DNA from various eukaryotic parasites associated with primates. In summary, these findings underscore the potential role of flies as potential vectors facilitating interactions between non-human primates, livestock, and humans in this ecologically diverse region.

In the third and final study of this thesis, I explored the potential of bulk fly pooling based destructive and non-destructive extraction schemes. I did this using flies captured across eight different sites in five African countries representing three different habitats and subsequently compared the findings with a previous study by Gogarten et al. (2020), which used the same flies but extracted destructively in smaller pools. While both of the bulk fly pool based extraction methods resulted in a significant reduction of cost and resources, the bulk non-destructive extraction method provided similarly effective to the more labour intensive smaller pool extraction methods used previously. This suggest that in some cases, particularly where pooling does not lose valuable information, as would be the case for some disease ecology investigations, the non-destructive bulk pooling method may be the most effective strategy.

In conclusion, the use of fly iDNA for biomonitoring represents a promising avenue for investigating disease ecology across various habitat types. Simultaneously, flies at human-wildlife interfaces may move between these habitats, thus highlighting their potential to serve as potential vectors for emerging infectious pathogens in the wild. And finally, it appears that the approach of bulk pooling fly bodies and subsequently employing a non-digestion buffer for DNA extraction has the potential to streamline the conventional fly extraction process, and has the potential to help bring these methods into more resource poor research laboratories. These innovations bring with them the potential to facilitate the widespread use of fly iDNA for terrestrial biomonitoring.

6 Zusammenfassung

Die Verwendung von Fliegen-iDNA zur Überwachung von Wildtieren und zur Erforschung der ökologischen Zusammenhänge von sylvatischem Milzbrand an Mensch-Wildtier-Grenzflächen

Die Überwachung von Wildtieren und Krankheiten spielt eine entscheidende Rolle bei der Erforschung unserer sich wandelnden Welt; nicht nur, um Erkenntnisse über die ökologische Dynamik und die Vorsorge gegen Pandemien zu gewinnen, sondern auch, um wirksame Strategien zur Bekämpfung des weltweiten Rückgangs der biologischen Vielfalt zu entwickeln, insbesondere in ressourcenarmen tropischen Hotspots der biologischen Vielfalt. Wirbellose eDNA-Quellen wie die iDNA von Fliegen haben ihren Wert als kosten- und ressourceneffizientes Biomonitoring-Instrument unter Beweis gestellt. Sie sind nicht nur wirksam bei der Erfassung der Artenvielfalt wildlebender Tiere und der Identifizierung wildlebender Krankheitserreger, sondern versprechen auch eine eingehende Erforschung verschiedener Aspekte der Ökologie wildlebender Tiere und Krankheitserreger. In Anbetracht der erfolgreichen Anwendung von Fliegen-iDNA in früheren Studien für beide Aspekte - den Nachweis von Tieren und Krankheitserregern - habe ich hier untersucht, ob Fliegen-iDNA Einblicke in die Krankheitsökologie entlang verschiedener Habitattypen geben kann. Außerdem untersuchte ich das Potenzial von Fliegen als Überträger in diesen verschiedenen Lebensräumen, insbesondere von Fliegen, die mit Primaten assoziiert sind, und ihr Potenzial, Krankheitserreger aus dem Urwald in vom Menschen genutzte Gebiete zu tragen. Schließlich untersuchte ich Möglichkeiten zur Optimierung bestehender iDNA-Analyseverfahren für Fliegen, um den Durchsatz bei der Probennahme zu maximieren und die Kosten zu senken, damit die biologische Vielfalt in größerem Maßstab bewertet werden kann.

In der ersten Studie wurde die iDNA von Fliegen auf molekularbiologischer Ebene analysiert, um die geografische Verbreitung des sylvatischen Milzbranderregers Bacillus cereus biovar anthracis (Bcbva) zu untersuchen, und es zeigten sich erhebliche Unterschiede beim Nachweis von Bcbva in den verschiedenen Lebensräumen vom Taï-Regenwald bis zu den nahe gelegenen Dörfern. Darüber hinaus ergab die Sequenzierung des gesamten Genoms von Bcbva aus positiven Fliegenpools deutliche Übereinstimmungen mit der genomischen Vielfalt von Bcbva, die durch jahrzehntelange Überwachungsaktivitäten beschrieben wurde. Es wurde auch eine

Metabarcodierung der iDNA von Fliegen durchgeführt, um die Zusammensetzung der Fliegen- und Säugetiergemeinschaften entlang dieses ökologischen Gradienten zu analysieren und mögliche Zusammenhänge mit dem Nachweis von Bcbva zu untersuchen. Die häufigste Entdeckung von Bcbva erfolgte in dem Gebiet, in dem auch die vielfältigste Säugetiergemeinschaft nachgewiesen wurde, nämlich im Wald. Im dörflichen Lebensraum, in dem die am geringsten diversifizierte Säugetiergemeinschaft nachgewiesen wurde, wurde kein Bcbva nachgewiesen, was auf eine starke Abhängigkeit von Bcbva vom Ökosystem des Regenwaldes schließen lässt. Die Studie zeigte jedoch auch, dass die größte Fliegengemeinschaft am Waldrand zu finden war, wo Fliegen sowohl aus anthropogenen als auch aus Waldbereichen stammten, was das potenzielle Risiko von Fliegen als Krankheitsüberträger zwischen den Wildtieren des Regenwaldes und den menschlichen Bewohnern in den umliegenden Dörfern verdeutlicht.

Daher habe ich in der zweiten Studie ein Markierungs-Wiederfang-Experiment durchgeführt, um die Bewegungsmuster von Fliegen zu untersuchen, die mit nicht-menschlichen Primaten in der Nähe des Kibale-Nationalparks in Uganda vergesellschaftet sind. Markierte Fliegen, die sich in Gruppen von nicht-menschlichen Primaten aufhielten, wurden anschließend in von Menschen genutzten Gebieten wieder eingefangen, wobei 19 markierte Fliegen mit einem Wiederfangaufwand von 28.615 Fliegen gefangen wurden. Die Anwendung des Metabarcoding auf die in nicht-menschlichen Primatengruppen gesammelten Fliegen enthüllte das Vorhandensein von DNA verschiedener eukaryotischer Parasiten, die mit Primaten assoziiert sind. Zusammenfassend unterstreichen diese Ergebnisse die potenzielle Rolle der Fliegen als potenzielle Überträger, die in dieser ökologisch vielfältigen Region Interaktionen zwischen nicht-menschlichen Primaten, Nutztieren und Menschen ermöglichen.

In der dritten und letzten Studie dieser Arbeit untersuchte ich das Potenzial von destruktiven und nicht-destruktiven Extraktionsverfahren auf der Basis von Fliegen-Pools. Dazu verwendete ich Fliegen, die an acht verschiedenen Standorten in fünf afrikanischen Ländern in drei verschiedenen Habitaten gefangen wurden, und verglich die Ergebnisse mit einer früheren Studie von Gogarten et al. (2020), in der dieselben Fliegen verwendet, aber destruktiv in kleineren Pools extrahiert wurden. Während beide Extraktionsmethoden auf der Grundlage von großen Fliegen-Pools zu einer erheblichen Verringerung der Kosten und Ressourcen führten, war die nicht-destruktive Extraktionsmethode ähnlich effektiv wie die zuvor verwendeten arbeitsintensiveren Extraktionsmethoden für kleinere Pools. Dies deutet darauf hin, dass in einigen Fällen, insbesondere wenn durch das Pooling keine wertvollen Informationen verloren gehen, wie dies

bei einigen krankheitsökologischen Untersuchungen der Fall ist, die nicht-destruktive Pooling-Methode die effektivste Strategie sein könnte.

Zusammenfassend lässt sich sagen, dass die Verwendung von Fliegen-iDNA für das Biomonitoring einen erfolgversprechenden Weg zur Untersuchung der Krankheitsökologie in verschiedenen Habitattypen darstellt. Gleichzeitig können sich Fliegen an den Schnittstellen zwischen Mensch und Tier zwischen diesen Lebensräumen bewegen, was ihr Potenzial als potenzielle Vektoren für neu auftretende infektiöse Erreger in der freien Natur verdeutlicht. Und schließlich hat es den Anschein, dass der gewählte Ansatz, Fliegenkörper in einem Pool zusammenzufassen und anschließend einen Puffer ohne Verdauung für die DNA-Extraktion zu verwenden, das Potenzial hat, den konventionellen Fliegenextraktionsprozess zu optimieren und dazu beizutragen, dass diese Methoden auch in ressourcenarmen Forschungslaboren eingesetzt werden können. Diese Verbesserungen haben das Potenzial, die weit verbreitete Anwendung von iDNA aus Fliegen für das terrestrische Biomonitoring zu erleichtern.

7 References

1. Adams C I M, Knapp M, Gemmell N J, Jeunen G J, Bunce M, Lamare M D, Taylor H R (2019): Beyond biodiversity: can environmental dna (edna) cut it as a population genetics tool? Genes (Basel), 10(3), 192. DOI: 10.3390/genes10030192.

2. Ahumada J A, Silva C E F, Gajapersad K, Hallam C, Hurtado J, Martin E, McWilliam A, Mugerwa B, O'Brien T, Rovero F, Sheil D, Spironello W R, Winarni N, Andelman S J (2011): Community structure and diversity of tropical forest mammals: data from a global camera trap network. Philos Trans R Soc Lond B Biol Sci., 366, 2703-11. DOI: 10.1098/rstb.2011.0115.

3. Aide T M, Hernández-Serna A, Campos-Cerqueira M, Acevedo-Charry O, Deichmann J L (2017): Species richness (of insects) drives the use of acoustic space in the tropics. Remote Sensing, 9, 1-12. DOI: 10.3390/rs9111096.

4. Alfano N, Dayaram A, Axtner J, Tsangaras K, Kampmann M L, Mohamed A, Wong S T, Gilbert M T P, Wilting A and Greenwood A D (2021): Non-invasive surveys of mammalian viruses using environmental DNA. Methods in Ecology and Evolution, 12, 1941-1952. DOI: 10.1111/2041-210X.13661.

5. Amstrup S C, McDonald T L, Manly B F J (2005): Handbook of capture-recapture analysis. Princeton University Press, Princeton.

6. Andersen K, Bird K L, Rasmussen M, Haile J, Breuning-Madsen H, Kjaer K H, Orlando L, Gilbert M T, Willerslev E (2012): Meta-barcoding of 'dirt' DNA from soil reflects vertebrate biodiversity. Mol Ecol., 8, 1966-79. DOI: 10.1111/j.1365-294X.2011.05261.x.

7. Andres K J, Lodge D M, Sethi S A, Andrés J (2023): Detecting and analysing intraspecific genetic variation with eDNA: From population genetics to species abundance. Mol Ecol., 4118-4132. DOI: 10.1111/mec.17031.

8. Antonation K S, Grützmacher K, Dupke S, Mabon P, Zimmermann F, Lankester F, Peller T, Feistner A, Todd A, Herbinger I, de Nys H M, Muyembe-Tamfun J J, Karhemere S, Wittig R M, Couacy-Hymann E, Grunow R, Calvignac-Spencer S, Corbett C R, Klee S R, Leendertz F H (2016): *Bacillus cereus* biovar *anthracis* causing anthrax in sub-saharan

africa-chromosomal monophyly and broad geographic distribution. PLoS Negl Trop Dis., 10, 1-14. DOI: 10.1371/journal.pntd.0004923.

9. Banjo A D, Lawal O A, Adeduji O O (2005): Bacteria and fungi isolated from housefly (*Musca domestica* L.) larvae. African Journal of Biotechnology, 4(8), 780-784. https://www.ajol.info/index.php/ajb/article/view/15181.

10. Bass D, Stentiford G D, Littlewood D T J, Hartikainen H (2015): Diverse applications of environmental dna methods in parasitology. Trends Parasitol., 31(10), 499-513. DOI: 10.1016/j.pt.2015.06.013.

11. Bitome-Essono P Y, Ollomo B, Arnathau C, Durand P, Mokoudoum N D, Yacka-Mouele L, Okouga A P, Boundenga L, Mve-Ondo B, Obame-Nkoghe J, Mbehang-Nguema P, Njiokou F, Makanga B, Wattier R, Ayala D, Ayala F J, Renaud F, Rougeron V, Bretagnolle F, Prugnolle F, Paupy C (2017): Tracking zoonotic pathogens using blood-sucking flies as 'flying syringes'. eLife 6,1-21. DOI: 10.7554/eLife.22069.

12. Boadella M, Gortazar C, Acevedo P, Carta T, Martín-Hernando M P, de la Fuente J, Vicente J (2011): Six recommendations for improving monitoring of diseases shared with wildlife: Examples regarding mycobacterial infections in Spain. European Journal of Wildlife Research, 57, 697-706. DOI: 10.1007/s10344-011-0550-x.

13. Bohmann K, Evans A, Gilbert M T P, Carvalho G R, Creer S, Knapp M, Yu D W, de Bruyn M (2014): Environmental DNA for wildlife biology and biodiversity monitoring. Trends in Ecology and Evolution, 29, 358-367. DOI: 10.1016/j.tree.2014.04.003.

14. Bohmann K, Schnell I B, Gilbert M (2013): When bugs reveal biodiversity. Mol Ecol., 22(4), 909-911. DOI: 10.1111/mec.12221.

15. Braack L E, De Vos V(1990): Feeding habits and flight range of blow-flies (*Chrysomyia* sp.) in relation to anthrax transmission in the Kruger National Park, South Africa. Onderstepoort J Vet Res., 57(2), 141-142. PMID: 2216348.

16. Brauer C L, Donovan T M, Mickey R M, Katz J and B. R. Mitchell (2016): A comparison of acoustic monitoring methods for common anurans of the northeastern United States. Wildlife Society Bulletin 40, 140-149. DOI: 10.1002/wsb.619.

17. Brownlie J, Peckham C, Waage J, Woolhouse M, Lyall C, Meagher L, Tait J, Baylis M, Nicoll A (2006): Foresight. Infectious Diseases: preparing for the future Future Threats. Office of Science and Innovation, London.

18. Butchart S H M, Walpole M, Collen B, Strien A V, Scharlemann J P W, Almond R E A, Baillie J E M, Bomhard B, Brown C, Bruno J, Carpenter K E, Carr G M, Chanson J, Chenery A M, Csirke J, Davidson N C, Dentener F, Foster M, Galli A, Galloway J N, Genovesi P, Gregory R D, Hockings M, Kapos V, Lamarque J F, Leverington F, Loh J, McGeoch M A, McRae L, Minasyan A, Morcillo M H, Oldfield T E E, Pauly D, Quader S, Revenga C, Sauer J R, Skolnik B, Spear D, Stanwell-Smith D, Stuart S N, Symes A, Tierney M, Tyrrell T D, Vié J C, Watson R (2010): Global biodiversity: Indicators of recent declines. Science, 328, 1164-1168. DOI:10.1126/science.1187512.

19. Calvignac-Spencer S, Leendertz F H, Gilbert M T P, Schubert G (2013a): An invertebrate stomach's view on vertebrate ecology: Certain invertebrates could be used as "vertebrate samplers" and deliver DNA-based information on many aspects of vertebrate ecology. BioEssays 35, 1004-1013. DOI: 10.1002/bies.201300060.

20. Calvignac-Spencer S, Merkel K, Kutzner N, Kühl H, Boesch C, Kappeler P M, Metzger S, Schubert G, Leendertz F H (2013b): Carrion fly-derived DNA as a tool for comprehensive and cost-effective assessment of mammalian biodiversity. Mol Ecol., 22, 915-924. DOI: 10.1111/mec.12183.

21. Campbell G, Kuehl H, Diarrassouba A, N'Goran P K, Boesch C (2011): Long-term research sites as refugia for threatened and over-harvested species. Biol Lett., 7, 723-6. DOI: 10.1098/rsbl.2011.0155.

22. Caravaggi A, Banks P B, Burton A C, Finlay C M V, Haswell P M, Hayward M W, Rowcliffe M J, Wood M D (2017): A review of camera trapping for conservation behaviour research, Remote Sensing in Ecology and Evolution , 3 (3),109-122. DOI:10.1002/rse2.

23. Carvalho C S, de Oliveira M E, Rodriguez-Castro K G, Saranholi B H, Galetti P M Jr. (2022): Efficiency of eDNA and iDNA in assessing vertebrate diversity and its abundance. Mol Ecol Resour. 22(4), 1262-1273. DOI: 10.1111/1755-0998.13543.

24. Chua K B (2003): Nipah virus outbreak in Malaysia. J Clin Virol., 26(3), 265-275. DOI: 10.1016/s1386-6532(02)00268-8.

25. Civitello D J, Cohen J, Fatima H, Halstead N T, Liriano J, McMahon T A, Ortega C N, Sauer E L, Sehgal T, Young S, Rohr J R (2015): Biodiversity inhibits parasites: Broad evidence for the dilution effect. Proc Natl Acad Sci U S A, 112(28), 8667-8671. DOI: 10.1073/pnas.1506279112.

26. Collett R A, Fisher D O (2017): Time-lapse camera trapping as an alternative to pitfall trapping for estimating activity of leaf litter arthropods. Ecol Evol.,7, 7527–7533. https://doi.org/10.1002/ece3.3275.

27. Cryan P M, Meteyer C U, Boyles J G, Blehert D S (2010): Wing pathology of white-nose syndrome in bats suggests life-threatening disruption of physiology. BMC Biol.,135, 1-8. DOI: 10.1186/1741-7007-8-135.

28. Danabalan R, Merkel K, Bærholm Schnell I, Arandjelovic M, Boesch C, Brazzola G, Dieguez P, Dupain J, Kambale-Vyalengerera M, Kühl H S, Hoffmann C (2023): Mammal mitogenomics from invertebrate-derived DNA. Environmental DNA. DOI: 10.1002/edn3.436.

29. Darras K, Batáry P, Furnas B J, Grass I, Mulyani Y A, Tscharntke T (2019): Autonomous sound recording outperforms human observation for sampling birds: a systematic map and user guide. Ecol Appl. 2019 Sep 29. DOI: 10.1002/eap.1954.

30. Daszak P, Cunningham A A, Hyatt A D (2000): Emerging infectious diseases of wildlife - Threats to biodiversity and human health. Science, 287, 443-449. DOI: 10.1126/science.287.5452.443.

31. Daszak, P, Cunningham A A, Hyatt A D (2001): Anthropogenic environmental change and the emergence of infectious diseases in wildlife. Acta tropica, 78(2), 103-116. DOI: 10.1016/S0001-706X(00)00179-0.

32. De Nys H M, Calvignac-Spencer S, Thiesen U, Boesch C, Wittig R M, Mundry R, Leendertz F H (2013): Age-related effects on malaria parasite infection in wild chimpanzees. Biol Lett., 9(4). DOI: 10.1098/rsbl.2012.1160.

33. Devaux C A, Mediannikov O, Medkour H and Raoult D (2019): Infectious disease risk across the growing human-non human primate interface: a review of the evidence. Frontiers in Public Health, 7, 1-22. DOI: 10.3389/fpubh.2019.00305.

34. Dirzo R, Young H S, Galetti M, Ceballos G, Isaac N J B, Collen B (2014): Defaunation in the anthropocene. Science, 345, 401-406. DOI: 10.1126/science.1251817.

35. Doherty T S, Glen A S, Nimmo D G, Ritchie E G, Dickman C R (2016): Invasive predators and global biodiversity loss. Proceedings of the National Academy of Sciences of the United States of America, 113, 11261-11265. DOI: 10.1073/pnas.1602480113.

36. Duggan M T, Groleau M F, Shealy E P, Self L S, Utter T E, Waller M M, Hall B C, Stone C G, Anderson L L, Mousseau T A (2021): An approach to rapid processing of camera trap images with minimal human input. Ecol Evol., 11(17), 12051-12063. DOI: 10.1002/ece3.7970.

37. Dupke S, Schubert G, Beudjé F, Barduhn A, Pauly M, Couacy-Hymann E, Grunow R, Akoua-Koffi C, Leendertz F H, Klee S R (2020): Serological evidence for human exposure to *Bacillus cereus* biovar *anthracis* in the villages around Taï National Park, Côte d'Ivoire. PLoS Negl Trop Dis., 14(5), e0008292. DOI: 10.1371/journal.pntd.0008292.

38. Farrell J A, Whitmore L, Duffy D J (2021): The promise and pitfalls of environmental DNA and RNA approaches for the monitoring of human and animal pathogens from aquatic sources. BioScience 71, 609-625. DOI: 10.1093/biosci/biab027.

39. Ficetola G F, Miaud C, Pompanon F, Taberlet P (2008): Species detection using environmental DNA from water samples. Biology Letters, 4,423-425. DOI: 10.1098/rsbl.2008.0118.

40. Förster M, Klimpel S, Mehlhorn H, Sievert K, Messler S, Pfeffer K (2007): Pilot study on synanthropic flies (e.g. Musca, Sarcophaga, Calliphora, Fannia, Lucilia, Stomoxys) as vectors of pathogenic microorganisms. Parasitol Res. 101(1), 243-6. DOI: 10.1007/s00436-007-0522-y.

41. Ganchev T, Potamitis I (2007): Automatic acoustic identification of singing insects. Bioacoustics, 16, 281-328. DOI: 10.1080/09524622.2007.9753582.

42. Gaynor K M, Hojnowski C E, Carter N H, Brashares J S (2018): The influence of human disturbance on wildlife nocturnality. Science, 360(6394), 1232-1235. DOI: 10.1126/science.aar7121.

43. Gilbert M T, Moore W, Melchior L, Worobey M (2007): DNA extraction from dry museum beetles without conferring external morphological damage. PLoS One. 2(3), e272. DOI: 10.1371/journal.pone.0000272.

44. Gilbert N A, Clare J D J, Stenglein J L, Zuckerberg B (2021): Abundance estimation of unmarked animals based on camera-trap data. Conserv Biol., 35(1):88-100. DOI: 10.1111/cobi.13517.

45. Gogarten J F, Düx A, Mubemba B, Pléh K, Hoffmann C, Mielke A, Müller-Tiburtius J, Sachse A, Wittig R M, Calvignac-Spencer S, Leendertz F H (2019): Tropical rainforest

flies carrying pathogens form stable associations with social nonhuman primates. Mol Ecol., 18, 4242-4258. DOI: 10.1111/mec.15145.

46. Gogarten J F, Hoffmann C, Arandjelovic M, Sachse A, Merkel K, Dieguez P, Agbor A, Angedakin S, Brazzola G, Jones S, Langergraber K E, Lee K, Marrocoli S, Murai M, Sommer V, Kühl H, Leendertz F H, Calvignac-Spencer S (2020): Fly-derived DNA and camera traps are complementary tools for assessing mammalian biodiversity. Environmental DNA 2, 63-76. DOI: 10.1002/edn3.46.

47. Gogarten J F, Jahan M, Calvignac-Spencer S, Chapman C A, Goldberg T L, Leendertz F H, Rothman J M (2022): The cost of living in larger primate groups includes higher fly densities. Ecohealth, 19(2), 290-298. DOI: 10.1007/s10393-022-01597-5.

48. Goldberg C S, Pilliod D S, Arkle R S, Waits L P (2011): Molecular detection of vertebrates in stream water: a demonstration using Rocky Mountain tailed frogs and Idaho giant salamanders. PLoS One. 6(7), e22746. DOI: 10.1371/journal.pone.0022746.

49. Graczyk T K, Knight R, Gilman R H, Cranfield M R (2001): The role of non-biting flies in the epidemiology of human infectious diseases. Microbes Infect., 3(3), 231-5. DOI: 10.1016/s1286-4579(01)01371-5.

50. Greenberg B (1971): Flies and disease, Vol. 1: Ecology, classification and biotic associations; flies and disease. Volume II. Biology and disease transmission. Princeton University Press.

51. Greenberg, B. (1973): Flies and disease, vol. II. Biology and transmission of disease, Princeton University Press, Princeton, NJ.

52. Grier J W (1982): Ban of DDT and subsequent recovery of reproduction in bald eagles. Science, 218(4578), 1232-1235. DOI: 10.1126/science.7146905.

53. Grubaugh N D, Sharma S, Krajacich B J, Iii L S F, Bolay F K, Ii J W D, Johnson W E, Ebel G D, Foy B D, Brackney D E (2015): Xenosurveillance : a novel mosquito-based approach for examining the human- pathogen landscape. PLoS Negl Trop Dis., 3, e0003628. DOI: 10.1371/journal.pntd.0003628.

54. Hebert P D, Cywinska A, Ball S L, de Waard J R (2003): Biological identifications through DNA barcodes. Proc Biol Sci., 270(1512), 313-321. DOI: 10.1098/rspb.2002.2218.

55. Hockings K J, Mubemba B, Avanzi C, Pleh K, Düx A, Bersacola E, Bessa J, Ramon M, Metzger S, Patrono L V, Jaffe J E, Benjak A, Bonneaud C, Busso P, Couacy-Hymann E,

Gado M, Gagneux S, Johnson R C, Kodio M, Lynton-Jenkins J, Morozova I, Mätz-Rensing K, Regalla A, Said A R, Schuenemann V J, Sow S O, Spencer J S, Ulrich M, Zoubi H, Cole S T, Wittig R M, Calvignac-Spencer S, Leendertz F H (2021): Leprosy in wild chimpanzees. Nature, 598(7882), 652-656. DOI: 10.1038/s41586-021-03968-4.

56. Hoekstra J M, Boucher T M, Ricketts T H, Roberts C (2005): Confronting a biome crisis: Global disparities of habitat loss and protection. Ecology Letters, 8, 23-29. DOI: 10.1111/j.1461-0248.2004.00686.x.

57. Hoffmann C, Merkel K, Sachse A, Rodríguez P, Leendertz F H, Calvignac-Spencer S (2018): Blow flies as urban wildlife sensors. Molecular Ecology Resources, 18, 502-510. DOI: 10.1111/1755-0998.12754.

58. Hoffmann C, Stockhausen M, Merkel K, Calvignac-Spencer S, Leendertz F H (2016): Assessing the feasibility of fly based surveillance of wildlife infectious diseases. Scientific Reports, 6, 1-9. DOI: 10.1038/srep37952.

59. Hoffmann C, Zimmermann F, Biek R, Kuehl H, Nowak K, Mundry R, Agbor A, Angedakin S, Arandjelovic M, Blankenburg A, Brazolla G, Corogenes K, Couacy-Hymann E, Deschner T, Dieguez P, Dierks K, Düx A, Dupke S, Eshuis H, Formenty P, Yuh Y G, Goedmakers A, Gogarten J F, Granjon A C, McGraw S, Grunow R, Hart J, Jones S, Junker J, Kiang J, Langergraber K, Lapuente J, Lee K, Leendertz S A, Léguillon F, Leinert V, Löhrich T, Marrocoli S, Mätz-Rensing K, Meier A, Merkel K, Metzger S, Murai M, Niedorf S, De Nys H, Sachse A, van Schijndel J, Thiesen U, Ton E, Wu D, Wieler L H, Boesch C, Klee S R, Wittig R M, Calvignac-Spencer S, Leendertz F H (2017): Persistent anthrax as a major driver of wildlife mortality in a tropical rainforest. Nature, 548(7665), 82-86. DOI: 10.1038/nature23309.

60. Huggins L G, Michaels C J, Cruickshank S M, Preziosi R F, Else K J (2017): A novel copro-diagnostic molecular method for qualitative detection and identification of parasitic nematodes in amphibians and reptiles. PLoS One. 12(9), e0185151. DOI: 10.1371/journal.pone.0185151.

61. Jahan M, Lagostina L, Gräßle T, Couacy-Hymann E, Kouadio L, Kouakou V K, Krou H A, Mossoun A M, Patrono L V, Pléh K and Steiner J A (2023): Fly iDNA suggests strict reliance of the causative agent of sylvatic anthrax on rainforest ecosystems. Environmental DNA, 1-12. DOI: 10.1002/edn3.401.

62. Jangra S, Ghosh A (2022): Rapid and zero-cost DNA extraction from soft-bodied insects for routine PCR-based applications. PLoS One. 17(7), e0271312. DOI: 10.1371/journal.pone.0271312.

63. Jerde C L, Mahon A R, Chadderton W L, Lodge D M (2011): “Sight-unseen” detection of rare aquatic species using environmental DNA. Conservation Letters, 4, 150-157. DOI: 10.1111/j.1755-263X.2010.00158.x.

64. Ji Y, Baker C C M, Popescu V D, Wang J, Wu C, Wang Z, Li Y, Wang L, Hua C, Yang Z, Yang C, Xu C C Y, Diana A, Wen Q, Pierce N E, Yu D W (2022): Measuring protected-area effectiveness using vertebrate distributions from leech iDNA. Nat Commun., 13(1), 1555. DOI: 10.1038/s41467-022-28778-8.

65. Jones K E, Patel N G, Levy M A, Storeygard A, Balk D, Gittleman J L, Daszak P (2008): Global trends in emerging infectious diseases. Nature, 451, 990-993. DOI: 10.1038/nature06536.

66. Junqueira A C M, Ratan A, Acerbi E, Drautz-Moses D I, Premkrishnan B N V, Costea P I, Linz B, Purbojati R W, Paulo D F, Gaultier N E, Subramanian P, Hasan N A, Colwell R R, Bork P, Azeredo-Espin A M L, Bryant D A, Schuster S C (2017): The microbiomes of blowflies and houseflies as bacterial transmission reservoirs. Scientific Reports, 7, 1-15. DOI: 10.1038/s41598-017-16353-x.

67. Keele B F, Jones J H, Terio K A, Estes J D, Rudicell R S, Wilson M L, Li Y, Learn G H, Beasley T M, Schumacher-Stankey J, Wroblewski E, Mosser A, Raphael J, Kamenya S, Lonsdorf E V, Travis D A, Mlengeya T, Kinsel M J, Else J G, Silvestri G, Goodall J, Sharp P M, Shaw G M, Pusey A E, Hahn B H (2009): Increased mortality and AIDS-like immunopathology in wild chimpanzees infected with SIVcpz. Nature 460, 515-519. DOI: 10.1038/nature08200.

68. Kent R J (2009): Molecular methods for arthropod bloodmeal identification and applications to ecological and vector-borne disease studies. Molecular Ecology Resources, 9, 4-18. DOI: 10.1111/j.1755-0998.2008.02469.x.

69. King H C, Murphy A, James P, Travis E, Porter D, Sawyer J, Cork J, Delahay R J, Gaze W, Courtenay O, Wellington E M (2015): Performance of a Noninvasive Test for Detecting Mycobacterium bovis Shedding in European Badger (*Meles meles*) Populations. J Clin Microbiol., 53(7), 2316-23. DOI: 10.1128/JCM.00762-15.

70. Knauf S, Raphael J, Mitjà O, Lejora I AV, Chuma I S, Batamuzi E K, Keyyu J D, Fyumagwa R, Lüert S, Godornes C, Liu H, Schwarz C, Šmajs D, Grange P, Zinner D, Roos C, Lukehart S A (2016): Isolation of Treponema DNA from Necrophagous Flies in a Natural Ecosystem. EBioMedicine., 11, 85-90. DOI: 10.1016/j.ebiom.2016.07.033.

71. Kocher A, de Thoisy B, Catzeflis F, Valière S, Bañuls A L, Murienne J (2017): iDNA screening: Disease vectors as vertebrate samplers. Mol Ecol. 22,6478-6486. DOI: 10.1111/mec.14362.

72. Köndgen S, Schenk S, Pauli G, Boesch C, Leendertz F H (2010): Noninvasive monitoring of respiratory viruses in wild chimpanzees. Ecohealth, 7(3), 332-341. DOI: 10.1007/s10393-010-0340-z.

73. Kuisma E, Olson S H, Cameron K N, Reed P E, Karesh W B, Ondzie A I, Akongo M J, Kaba S D, Fischer R J, Seifert S N, Muñoz-Fontela C, Becker-Ziaja B, Escudero-Pérez B, Goma-Nkoua C, Munster V J, Mombouli J V (2019): Long-term wildlife mortality surveillance in northern Congo: a model for the detection of Ebola virus disease epizootics. Philos Trans R Soc Lond B Biol Sci., 374, 1-8. DOI: 10.1098/rstb.2018.0339.

74. Lampa S, Mihoub J B, Gruber B, Klenke R, Henle K (2015): Non-invasive genetic mark-recapture as a means to study population sizes and marking behaviour of the elusive eurasian otter (*Lutra lutra*). PLoS One, 14, 10(5). DOI: 10.1371/journal.pone.0125684.

75. Laurance W F, Sayer J, Cassman K G (2014): Agricultural expansion and its impacts on tropical nature. Trends Ecol Evol., 29, 107-116. DOI: 10.1016/j.tree.2013.12.001.

76. Lee P S, Gan H M, Clements G R, Wilson J J (2016): Field calibration of blowfly-derived DNA against traditional methods for assessing mammal diversity in tropical forests. Genome, 59(11),1008-1022. DOI: 10.1139/gen-2015-0193.

77. Lee P S, Sing K W, Wilson J J (2015): Reading mammal diversity from flies: the persistence period of amplifiable mammal mtDNA in blowfly guts (*Chrysomya megacephala*) and a new DNA mini-barcode target. PLoS One.10 (4), e0123871. DOI: 10.1371/journal.pone.0123871.

78. Leendertz F H, Boesch C, Ellerbrok H, Rietschel W, Couacy-Hymann E, Pauli G (2004): Non-invasive testing reveals a high prevalence of simian T-lymphotropic virus type 1 antibodies in wild adult chimpanzees of the Taï National Park, Côte d'Ivoire. Journal of General Virology 85, 3305-3312. DOI: 10.1099/vir.0.80052-0.

79. Leendertz F H, Pauli G, Maetz-Rensing K, Boardman W, Nunn C, Ellerbrok H, Jensen S A, Junglen S, Christophe B (2006): Pathogens as drivers of population declines: The importance of systematic monitoring in great apes and other threatened mammals. Biological Conservation, 131, 325-337. DOI: 10.1016/j.biocon.2006.05.002.

80. Leroy E M, Rouquet P, Formenty P, Souquière S, Kilbourne A, Froment J M, Bermejo M, Smit S, Karesh W, Swanepoel R, Zaki S R, Rollin P E (2004): Multiple ebola virus transmission events and rapid decline of central african wildlife. Science 303, 387-390. DOI: 10.1126/science.1092528.

81. Levins R, Awerbuch T, Brinkmann U, Eckardt I, Epstein P, Makhoul N, Possas C A D, Puccia C, Spielman A, Mary E, Wilson M E (1994): The emergence of new diseases. American Scientist, 82: 5-5. http://www.jstor.org/stable/29775101

82. Lindenmayer D B, Lavery T, Scheele B C (2022): Why we need to invest in large-scale, long-term monitoring programs in landscape ecology and conservation biology. Curr Landscape Ecol Rep., 7, 137–146. https://doi.org/10.1007/s40823-022-00079-2

83. Liu W, Worobey M, Li Y, Keele B F, Bibollet-Ruche F, Guo Y, Goepfert P A, Santiago M L, Ndjango J B, Neel C, Clifford S L, Sanz C, Kamenya S, Wilson M L, Pusey A E, Gross-Camp N, Boesch C, Smith V, Zamma K, Huffman M A, Mitani J C, Watts D P, Peeters M, Shaw G M, Switzer W M, Sharp P M, Hahn B H (2008): Molecular ecology and natural history of simian foamy virus infection in wild-living chimpanzees. PLoS Pathog., 4, 4(7). DOI: 10.1371/journal.ppat.1000097.

84. Lukacs P M, Burnham K P(2005): Review of capture-recapture methods applicable to noninvasive genetic sampling. Molecular Ecology, 14, 3909-3919. DOI: 10.1111/j.1365-294X.2005.02717.x.

85. Lynggaard C, Nielsen M, Santos-Bay L, Gastauer M, Oliveira G, Bohmann K (2019): Vertebrate diversity revealed by metabarcoding of bulk arthropod samples from tropical forests. Environmental DNA, 1, 329-341. DOI: 10.1002/edn3.34.

86. Makuwa M, Souquière S, Clifford S L, Mouinga-Ondeme A, Bawe-Johnson M, Wickings E J, Latour S, Simon F, Roques P (2005): Identification of hepatitis B virus genome in faecal sample from wild living chimpanzee (Pan troglodytes troglodytes) in Gabon. J Clin Virol., 34 (Suppl 1), S83-8. DOI: 10.1016/s1386-6532(05)80016-2.

87. Mielke A, Zuberbühler K (2013): A method for automated individual, species and call type recognition in free-ranging animals. Animal Behaviour 86, 475-482. DOI: 10.1016/j.anbehav.2013.04.017.

88. Mombo I M, Fritz M, Becquart P, Liegeois F, Elguero E, Boundenga L, Mebaley T N, Prugnolle F, Maganga G D, Leroy E M (2020): Detection of ebola virus antibodies in fecal samples of great apes in Gabon. Viruses, 12, 1-10. DOI: 10.3390/v12121347.

89. Morin D J, Kelly M J, Waits L P (2016): Monitoring coyote population dynamics with fecal DNA and spatial capture–recapture. Journal of Wildlife Management, 80, 824-836. DOI: 10.1002/jwmg.21080.

90. Morse S S (1995): Factors in the emergence of infectious diseases. Emerg Infect Dis., 1, 7-15. DOI: 10.3201/eid0101.950102.

91. Nguyen, H D D, Nansen C (2018): Edge-biased distributions of insects. A review. agronomy for sustainable development, 38. DOI: 10.1007/s13593-018-0488-4.

92. Nikolin V M, Olarte-Castillo X A, Osterrieder N, Hofer H, Dubovi E, Mazzoni C J, Brunner E, Goller K V, Fyumagwa R D, Moehlman P D, Thierer D, East M L (2017): Canine distemper virus in the Serengeti ecosystem: molecular adaptation to different carnivore species. Mol Ecol., 26(7), 2111-2130. DOI: 10.1111/mec.13902.

93. Norouzzadeh M S, Nguyen A, Kosmala M, Swanson A, Palmer M S, Packer C, Clune J (2018): Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. Proceedings of the National Academy of Sciences of the United States of America 115, E5716-E5725. DOI: 10.1073/pnas.1719367115.

94. O'Brien T G, M F Kinnaird (2008): A picture is worth a thousand words: The application of camera trapping to the study of birds. Bird Conservation International 18, S144-S162. DOI: 10.1017/S0959270908000348.

95. Ogutu J O, Owen-Smith N, Piepho H P, Said M Y (2011): Continuing wildlife population declines and range contraction in the Mara region of Kenya during 1977-2009. Journal of Zoology, 285, 99-109. DOI: 10.1111/j.1469-7998.2011.00818.x.

96. Palsbøll P J, Allen J, Bérubé M, Clapham P J, Feddersen T P, Hammond P S, Hudson R R, Jørgensen H, Katona S, Larsen A H, Larsen F, Lien J, Mattila D K, Sigurjónsson J, Sears R, Smith T, Sponer R, Stevick P, Oien N (1997): Genetic tagging of humpback whales. Nature, 388(6644), 767-769. DOI: 10.1038/42005.

97. Park R, Dzialo M C, Spaepen S, Nsabimana D, Gielens K, Devriese H, Crauwels S, Tito R Y, Raes J, Lievens B, Verstrepen K J (2019): Microbial communities of the house fly *Musca domestica* vary with geographical location and habitat. Microbiome. 7(1), 147. DOI: 10.1186/s40168-019-0748-9.

98. Patrono L V, Pléh K, Samuni L, Ulrich M, Röthemeier C, Sachse A, Muschter S, Nitsche A, Couacy-Hymann E, Boesch C, Wittig R M (2020): Monkeypox virus emergence in wild chimpanzees reveals distinct clinical outcomes and viral diversity. Nat Microbiol., 5(7), 955-965. DOI: 10.1038/s41564-020-0706-0.

99. Piaggio A J, Engeman R M, Hopken M W, Humphrey J S, Keacher K L, Bruce W E, Avery M L (2014): Detecting an elusive invasive species: a diagnostic PCR to detect Burmese python in Florida waters and an assessment of persistence of environmental DNA. Mol Ecol Resour., 14(2), 374-380. DOI: 10.1111/1755-0998.12180.

100. Piret J, Boivin G (2021): Pandemics throughout history. Frontiers in Microbiology, 11. DOI: 10.3389/fmicb.2020.631736.

101. Plumptre A J (2000): Monitoring mammal populations with line transect techniques in African forests. Journal of Applied Ecology 37, 356-368. DOI: 10.1046/j.1365-2664.2000.00499.x.

102. Ramirez J I, Zwerts J A, van Kuijk M, Iacobelli P, Li X, Herdoiza N, Jansen P A (2021): Density dependence of daily activity in three ungulate species. Ecol Evol., 11(12), 7390-7398. DOI: 10.1002/ece3.7570.

103. Rees H C, Maddison B C, Middleditch D J, Patmore J R M, Gough K C (2014): The detection of aquatic animal species using environmental DNA- a review of eDNA as a survey tool in ecology. Journal of Applied Ecology, 51,1450-1459. DOI: 10.1111/1365-2664.12306.

104. Refisch J, Koné I (2005): Impact of commercial hunting on monkey populations in the Taï region, Côte d'Ivoire. Biotropica, 37, 136-144. DOI: 10.1111/j.1744-7429.2005.03174.x.

105. Richards C S, Rowlinson C C, Cuttiford L, Grimsley R, Hall M J (2013): Decomposed liver has a significantly adverse affect on the development rate of the blowfly Calliphora vicina. Int J Legal Med., 127(1), 259-262. DOI: 10.1007/s00414-012-0697-3.

106. Richardson E, Nimmo D G, Avitabile S, Tworkowski L, Watson S J, Welbourne D, Leonard S W (2018): Camera traps and pitfalls: an evaluation of two methods for surveying reptiles in a semiarid ecosystem. Wildlife Research, 44(8), 637-647. https://doi.org/10.1071/WR16048.

107. Rodgers T W, Mock K E (2015): Drinking water as a source of environmental DNA for the detection of terrestrial wildlife species. Conservation Genetics Resources 7, 693-696. DOI: 10.1007/s12686-015-0478-7.

108. Rodgers T W, Xu C C Y, Giacalone J, Kapheim K M, Saltonstall K, Vargas M, Yu D W, Somervuo P, McMillan W O, Jansen P A (2017): Carrion fly-derived DNA metabarcoding is an effective tool for mammal surveys: Evidence from a known tropical mammal community. Molecular Ecology Resources 17, 133-145. DOI: 10.1111/1755-0998.12701.

109. Romero-Alvarez D, Peterson A T, Salzer J S, Pittiglio C, Shadomy S, Traxler R, Vieira A R, Bower W A, Walke H, Campbell L P(2020): Potential distributions of *Bacillus anthracis* and *Bacillus cereus* biovar *anthracis* causing anthrax in Africa. PLoS Negl Trop Dis., 14, 1-20. DOI: 10.1371/journal.pntd.0008131.

110. Rovero F, Tobler M, Sanderson J (2010): Camera trapping for inventorying terrestrial vertebrates. Manual on field recording techniques and protocols for all taxa biodiversity inventories and monitoring. The Belgian National Focal Point to the Global Taxonomy Initiative, 8, 100-128.

111. Russo D, Voigt C C (2016): The use of automated identification of bat echolocation calls in acoustic monitoring: A cautionary note for a sound analysis. Ecological Indicators, 66, 598-602. DOI: 10.1016/j.ecolind.2016.02.036.

112. Ryser-Degiorgis M P (2013): Wildlife health investigations: needs, challenges and recommendations. BMC Vet Res. 9, 223. DOI: 10.1186/1746-6148-9-223.

113. Santiago M L, Rodenburg C M, Kamenya S, Bibollet-Ruche F, Gao F, Bailes E, Meleth S, Soong S J, Kilby J M, Moldoveanu Z, Fahey B, Muller M N, Ayouba A, Nerrienet E, McClure H M, Heeney J L, Pusey A E, Collins D A, Boesch C, Wrangham R W, Goodall J, Sharp P M, Shaw G M, Hahn B H (2002): SIVcpz in wild chimpanzees. Science, 295, 465-465. DOI: 10.1126/science.295.5554.465.

114. Schaumburg F, Mugisha L, Kappeller P, Fichtel C, Köck R, Köndgen S, Becker K, Boesch C, Peters G, Leendertz F (2013): Evaluation of non-invasive biological samples to monitor

Staphylococcus aureus colonization in great apes and lemurs. PLoS One, 8, 6-11, DOI: 10.1371/journal.pone.0078046.

115. Scheele B C, Pasmans F, Skerratt L F, Berger L, Martel A, Beukema W, Acevedo A A, Burrowes P A, Carvalho T, Catenazzi A, De la Riva I, Fisher M C, Flechas S V, Foster C N, Frías-Álvarez P, Garner T W J, Gratwicke B, Guayasamin J M, Hirschfeld M, Kolby J E, Kosch T A, La Marca E, Lindenmayer D B, Lips K R, Longo A V, Maneyro R, McDonald C A, Mendelson J 3rd, Palacios-Rodriguez P, Parra-Olea G, Richards-Zawacki C L, Rödel M O, Rovito S M, Soto-Azat C, Toledo L F, Voyles J, Weldon C, Whitfield S M, Wilkinson M, Zamudio K R, Canessa S (2019): Amphibian fungal panzootic causes catastrophic and ongoing loss of biodiversity. Science, 363(6434), 1459-1463. DOI: 10.1126/science.aav0379.

116. Schmeller D S, Courchamp F, Killeen G (2020): Biodiversity loss, emerging pathogens and human health risks. Biodivers Conserv., 29, 3095-3102. DOI: 10.1007/s10531-020-02021-6.

117. Schnell I B, Thomsen P F, Wilkinson N, Rasmussen M, Jensen L R D, Willerslev E, Bertelsen M. F, Gilbert M T P (2012): Screening mammal biodiversity using DNA from leeches. Current Biology 22, R262-R263. DOI: 10.1016/j.cub.2012.02.058.

118. Schubert G, Stockhausen M, Hoffmann C, Merkel K, Vigilant L, Leendertz F H, Calvignac-Spencer S (2015): Targeted detection of mammalian species using carrion fly-derived DNA. Mol Ecol Resour., 15(2), 285-94. DOI: 10.1111/1755-0998.12306.

119. Scott Brandes T (2008): Automated sound recording and analysis techniques for bird surveys and conservation. Bird Conservation International 18, S163-S173. DOI: 10.1017/S0959270908000415.

120. Seeber P A, Soilemetzidou S E, East M L, Walzer C, Greenwood A D (2017): Equine behavioral enrichment toys as tools for non-invasive recovery of viral and host DNA. Zoo Biol., 36(5), 341-344. DOI: 10.1002/zoo.21380.

121. Snaddon J, Petrokofsky G, Jepson P, Willis K J (2013): Biodiversity technologies: Tools as change agents. Biology Letters, 9, 2012-2015. DOI: 10.1098/rsbl.2012.1029.

122. Soltis J, Leong K, Savage A (2005): African elephant vocal communication II: Rumble variation reflects the individual identity and emotional state of callers. Animal Behaviour, 70, 589-599. DOI: 10.1016/j.anbehav.2004.11.016.

123. Srivathsan A, Loh R K, Ong E J, Lee L, Ang Y, Kutty S N, Meier R (2022): Network analysis with either Illumina or MinION reveals that detecting vertebrate species requires metabarcoding of iDNA from a diverse fly community. Mol Ecol., 1-18. DOI: 10.1111/mec.16767.

124. Stadtländer C T K H (2015): One Health: People, animals, and the environment. Infection ecology & epidemiology, 5, 30514-30514. DOI: 10.3402/iee.v5.30514.

125. Stat M, Huggett M J, Bernasconi R, Dibattista J D, Berry T E, Newman S J, Harvey E S, Bunce M (2017): Ecosystem biomonitoring with eDNA: Metabarcoding across the tree of life in a tropical marine environment. Scientific Reports, 7, 1-11. DOI: 10.1038/s41598-017-12501-5.

126. Stat M, John J, DiBattista J D, Newman S J, Bunce M, Harvey E S (2019): Combined use of eDNA metabarcoding and video surveillance for the assessment of fish biodiversity. Conservation Biology, 33, 196-205. DOI: 10.1111/cobi.13183.

127. Sugai L S M, Silva T S F, Ribeiro J W, Llusia D (2019): Terrestrial Passive Acoustic Monitoring: Review and Perspectives. BioScience 69, 5-11. DOI: 10.1093/biosci/biy147.

128. Tabak M A, Norouzzadeh M S, Wolfson D W, Sweeney S J, Vercauteren K C, Snow N P, Halseth J M, Di Salvo P A, Lewis J S, White M D, Teton B, Beasley J C, Schlichting P E, Boughton R K, Wight B, Newkirk E S, Ivan J S, Odell E A, Brook R K, Lukacs P M, Moeller A K, Mandeville E G, Clune J, Miller R S (2019): Machine learning to classify animal species in camera trap images: Applications in ecology. Methods in Ecology and Evolution, 10, 585-590. DOI: 10.1111/2041-210X.13120.

129. Taberlet P, Bonin A, Zinger L, Coissac E (2018): Environmental DNA: For biodiversity research and monitoring. Oxford University Press.

130. Taberlet P, Coissac E, Hajibabaei M, Rieseberg L H (2012a): Environmental DNA. Mol Ecol., 8, 1789-93. DOI: 10.1111/j.1365-294X.2012.05542.x.

131. Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E (2012b): Towards next-generation biodiversity assessment using DNA metabarcoding. Mol Ecol., 8, 2045-50. DOI: 10.1111/j.1365-294X.2012.05470.x.

132. Taberlet P, Waits L P, Luikart G (1999): Noninvasive genetic sampling: look before you leap. Trends Ecol Evol., 14(8), 323-327. DOI: 10.1016/s0169-5347(99)01637-7.

133. Thomsen P F, Kielgast J, Iversen L L, Møller P R, Rasmussen M, Willerslev E (2012a): Detection of a Diverse Marine Fish Fauna Using Environmental DNA from Seawater Samples. PLoS ONE 7, 1-9. DOI: 10.1371/journal.pone.0041732.

134. Thomsen P F, Kielgast J, Iversen L L, Wiuf C, Rasmussen M, Gilbert M T P, Orlando L, Willerslev E (2012b): Monitoring endangered freshwater biodiversity using environmental DNA. Mol Ecol., 21, 2565-2573. DOI: 10.1111/j.1365-294X.2011.05418.x.

135. Troudet J, Grandcolas P, Blin A, Vignes-Lebbe R, Legendre F (2017): Taxonomic bias in biodiversity data and societal preferences. Scientific Reports, 7, 1-14. DOI: 10.1038/s41598-017-09084-6.

136. Valentini A, Taberlet P, Miaud C, Civade R, Herder J, Thomsen P F, Bellemain E, Besnard A, Coissac E, Boyer F, Gaboriaud C, Jean P, Poulet N, Roset N, Copp G H, Geniez P, Pont D, Argillier C, BauDOIn J-M, Peroux T, Crivelli A J, Olivier A, Acqueberge M, Le Brun M, Møller P R, Willerslev E, Dejean T (2016): Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. Mol. Ecol., 25, 929-942. https://DOI.org/10.1111/mec.13428.

137. Wearn O R, Glover-Kapfer P (2019): Snap happy: camera traps are an effective sampling tool when compared with alternative methods. R Soc Open Sci., 6(3), 181748. DOI: 10.1098/rsos.181748.

138. Wilcove D S, Giam X, Edwards D P, Fisher B, Koh L P (2013): Navjot's nightmare revisited: Logging, agriculture, and biodiversity in Southeast Asia. Trends Ecol Evol., 28, 531-540. DOI: 10.1016/j.tree.2013.04.005.

139. Woolhouse M E J (2002): Population biology of emerging and re-emerging pathogens. Trends in Microbiology, 10, 3-7. DOI: 10.1016/S0966-842X (02)02428-9.

140. Wrege P H, Rowland E D, Keen S, Shiu Y (2017): Acoustic monitoring for conservation in tropical forests: examples from forest elephants. Methods in Ecology and Evolution, 8, 1292-1301. DOI: 10.1111/2041-210X.12730.

141. Wright A M, Beres S B, Consamus E N, Long S W, Flores A R, Barrios R, Richter G S, Oh S Y, Garufi G, Maier H, Drews A L, Stockbauer K E, Cernoch P, Schneewind O, Olsen R J, Musser J M (2011): Rapidly progressive, fatal, inhalation anthrax-like infection in a human: case report, pathogen genome sequencing, pathology, and coordinated response. Arch Pathol Lab Med., 135(11), 1447-59. DOI: 10.5858/2011-0362-SAIR.1.

142. WWF (2022): Living Planet Report 2022 – Building a nature-positive society. Almond R E A, Grooten M, Juffe Bignoli, D. & Petersen, T. (Eds). WWF, Gland, Switzerland.

143. Yoccoz N G, Bråthen K A, Gielly L, Haile J, Edwards M E, Goslar T, Von Stedingk H, Brysting A K, Coissac E, Pompanon F, Sønstebø J H, Miquel C, Valentini A, De Bello F, Chave J, Thuiller W, Wincker P, Cruaud C, Gavory F, Rasmussen M, Gilbert M T, Orlando L, Brochmann C, Willerslev E, Taberlet P (2012): DNA from soil mirrors plant taxonomic and growth form diversity. Mol Ecol., (15), 3647-55. DOI: 10.1111/j.1365-294X.2012.05545.x.

144. Zwerts, J A, Stephenson P J, Maisels F, Rowcliffe M, Astaras C, Jansen P A, van der Waarde J, Sterck L E H M, Verweij P A, Bruce T, Brittain S, van Kuijk M (2021): Methods for wildlife monitoring in tropical forests: Comparing human observations, camera traps, and passive acoustic sensors. Conservation Science and Practice 3, 1-19. DOI: 10.1111/csp2.568.

8 List of Publications

Research articles

Jahan M, Lagostina L, Gräßle T, Couacy-Hymann E, Kouadio L, Kouakou V K, Krou H A, Mossoun A M, Patrono L V, Pléh K, Steiner J A, Yves N, Leendertz F H, Calvignac-Spencer S, & Gogarten J F (2023): Fly iDNA suggests strict reliance of the causative agent of sylvatic anthrax on rainforest ecosystems. Environmental DNA, 00, 1–12. https://doi.org/10.1002/edn3.401

Jahan M, Calvignac-Spencer S, Chapman C A, Kalbitzer U, Leendertz F H, Omeja P A, Sarkar D, Ulrich M, Gogarten J F (2022): The movement of pathogen carrying flies at the human–wildlife interface. EcoHealth, 19(4), 450-7, https://doi.org/10.1007/s10393-022-01621-8

Gogarten J F, Jahan M, Calvignac-Spencer S, Chapman C A, Goldberg T L, Leendertz F H, Rothman J M (2022): The cost of living in larger primate groups includes higher fly densities. EcoHealth, *19*(2), 290-298. https://doi.org/10.1007/s10393-022-01597-5

Poster presentations

Jahan M, Gogarten JF, Gräßle T, Lagostina L, Leonce K, Sébastien Calvignac-Spencer S, Leendertz FH. Fly iDNA shows reliance of sylvatic anthrax in rainforest ecosystem. Zoonoses 2022- International Symposium on Zoonoses Research, Berlin, (2022)

Jahan M, Gogarten JF, Calvignac-Spencer S. Exploring pathogen caring primate associated flies movement using bulk fly leg pools. Robert Koch Doctoral Students Symposium (RoKoCon), Berlin (2022)

Jahan M, Gogarten JF, Calvignac-Spencer S, Leendertz FH. Investigating host & pathogen diversity in large ecosystems using flies. Zoonoses meeting 2022, Hannover, (2022)

9 Acknowledgements

First I would like to express my utmost gratitude to my supervisor, Professor Dr. Fabian Leendertz for choosing me as one of his PhD students, providing me with the chance to work with this remarkable dataset, and for his innovative thinking that served as the foundation for numerous interesting projects. I would also acknowledge the outstanding contribution of another member of my supervisory team in my roller coaster PhD journey, my mentor, Professor Dr. Sébastien Calvignac-Spencer, who played a pivotal role in reigniting my research passion after the challenging period of COVID, when I had nearly given up hope of continuing my PhD journey. During that time, he unwaveringly stood by my side, offering staunch support and boundless patience. He not only imparted fundamental knowledge and insights into the latest molecular techniques available, but also encouraged me to venture into unexplored territories, demonstrating that fear should not deter us. 'Molecular biology is robust' is now a key component of how I approach my experiments, and for this I am deeply grateful. Most importantly, he consistently bolstered my self-confidence, reminding me of my own capabilities. Further, I would like to thank Prof. Dr. Marcus Fulde for supervising this work, for his valuable feedback and responsiveness and encouragement during our supervisory meetings.

I wish to commence my letter of appreciation by acknowledging one of my esteemed colleagues from our research team, Dr. Jan F. Gogarten. He is not only exceptionally talented but also one of the most humble individuals I have ever had the privilege of knowing. Throughout my entire PhD journey, he has served as an exemplary mentor, guide, and a steady support system, akin to a guardian angel.

I cannot overstate the significance of his support and guidance in enabling me to complete my PhD within a relatively short timeframe. When I initially embarked on my PhD journey, I felt utterly lost during the first few months. It was Dr. Gogarten who provided me with invaluable direction on how to gain a clear understanding of the entire research narrative. He guided me on how to organize my work effectively, ensuring that I remained on the right track and on schedule. Additionally, he generously involved me in two captivating projects that marked the true beginning of my PhD expedition. I also want to express my gratitude to him for sharing his profound passion for science, his innovative ideas, and for consistently being the most outstanding brainstorming partner. Without his substantial support, my rapid progress in thesis writing would not have been possible. It is impossible to adequately convey the extent of his contribution to my PhD journey,

and all I can say that I just have immense amount of respect and gratefulness for everything he has done for me and felt so privileged to work with him as a disciple.

I started my PhD journey right before the pandemic COVID happened to the world and shattered all of our normal living. My whole laboratory experiment was hugely interrupted by two consecutive lock down periods. All of this made managing the ambiguous lab-work I had planned after the pandemic extremely challenging and it was only possible with the constant support and sincere academic cooperation I got from Markus Ulrich, a fellow PhD student, an admirer of science and a long term member of this team, is an extremely brilliant and knowledgeable person who helped me a lot by showing great patience and sharing his skills. He provide all the information I needed to know about protocols and using machines and also guided me on my journey of learning bioinformatics and statistics. Moreover, I am forever indebted to Marcus for his invaluable assistance during the transition from Berlin to Greifswald, while our entire team relocated to the Helmholtz Institute for One Health.

From the core of my heart, I also thank to the lab technicians at the RKI, Andreas Sachse, Kevin Merkel and Caroline Röthemeier (Caro), who taught me how to organize and work in such a big research lab, which I had never experienced in my country. In particular, Andreas and Kevin who were never tired of answering my questions and helping me when something went wrong in the lab (which not only happened once). My special thanks to Andreas to help me to get familiarize with Berlin life by taking care of flat renting and organizing basic things for my living in completely unfamiliar world.

I want to express my immense gratitude to two of my wonderful colleagues and dear friends, Yanthe Nobel and Hattie Herridge. Even though Yanthe dedicated a substantial amount of her time to her research in Africa, whenever she was in Germany, she consistently offered constant encouragement and motivation. This was especially significant during the tough period when I received the news of my mother's endometrial carcinoma diagnosis in 2020. Her presence served as a continuous wellspring of strength, and she has been a great friend throughout this entire journey.

I also want to acknowledge Hattie, who joined our group after I had started my PhD, but quickly became a cherished friend. Her charismatic charm and vibrant personality endeared her to me from the very beginning. Throughout the past few years, as I grappled with extensive laboratory work, Hattie proved to be an invaluable colleague and a steadfast friend, always ready to cheer me on.

My heartfelt thanks to all other P3 members for their great support whenever needed and for creating a fun and friendly P3 atmosphere: Livia V. Patrono, Lorenzo Lagostina and Ariane Düx (best supportive post-Docs) and Our PhD team: Kamilla Pléh (who is very cooperative, sent me mask all the way from Hungary during the first lock down), Moritz Jochum (very dear friend), Jasmine Schlotterbeck, Olivia Demov, Jenny Jaffe and Khaled Ferchichi. I would like to specially thank Tobias Gräßle for introducing me to the PCR track at the beginning of my time at the RKI and also for his help in guiding me during my efforts in Bcbva whole genome sequencing; from the lab through the bioinformatics analyses.

My special thanks goes to Milen Sterev and Robin König (best supportive HiWi) who helped me a lot through sorting out the flies in the lab. In addition, I would like to appreciate all the efforts of my beloved colleagues from Biodiversa project, the Kibale Fish and Monkey Project and the PanAF fly project for collecting such valuable data and bring all of these flies from different parts of Africa to Germany. I have never been to Africa, but I hope our paths will cross and that I can thank you in person. Without their efforts and knowledge of the local environments, I would have never have been able to conduct all of these thesis projects. I also express my gratitude to all the members of the HIOH team for their tremendous support and warm hospitality.

I also would like to thank MF2 at the Robert Koch Institute, who gave me the opportunity to use their BSL3 laboratory for culturing Bcbva and all of the administrators in Berlin and Greifswald who have helped make my time in Germany such a resounding success, despite all of the challenges that a pandemic brings. For her help comments on some sections of this thesis, I am particularly grateful to Stephanie Markert for her patience and time.

Finally, I would like to extend my heartfelt gratitude to my parents and my beloved siblings. As parents from a third-world country, they faced immense challenges in supporting all three of children in pursuing higher education abroad. However, they never hesitated in their commitment to breaking social taboos and ensuring that we received the right education to fulfil our dreams. I would also like to express my gratitude to my sister, Noushin Jahan, and my younger brother, Sakibur Rahman. They have been a constant source of inspiration throughout this roller coaster journey. Their exceptional understanding and kindness during my mother's treatment, when I could not return to my country due to lockdown restrictions, meant the world to me.

For four years of funding I sincerely thank Bangabandhu Science and Technology fellowship trust, government of the People’s Republic of Bangladesh for offering me this prestigious scholarship. I hope to bring the skills and knowledge I have gained through this PhD project with me when I return and am excited to see where the next chapters of my research career take me.

10 Funding and Conflict of Interest

Funding Sources

The thesis work was supported by Bangabandhu Science and Technology fellowship trust, Ministry of Science and Technology, Government of the People's Republic of Bangladesh.

Conflict of Interest

Within the scope of this work, there are no conflict of interest due to third-party donations.

11 Declaration of Independence

I hereby confirm that the present dissertation has been independently prepared by me in its entirety, and that the resources used have been fully acknowledged. Publications of parts of the present dissertation are referenced. Furthermore, I declare that I have not previously submitted an application for a doctoral degree or initiated a doctoral procedure elsewhere.

Berlin, 01.12.2023
Mueena Jahan

Druck: mbvberlin mensch und buch verlag